



UNIVERSIDADE NOVA DE LISBOA

**DIVERSIDADE GENÉTICA E RESISTÊNCIA AOS
ANTI-RETROVIRAIS INIBIDORES ENZIMÁTICOS DE
VÍRUS DA IMUNODEFICIÊNCIA HUMANA TIPO 1 (HIV-1)
DE UMA POPULAÇÃO DE TOXICODEPENDENTES POR
VIA ENDOVENOSA DA GRANDE LISBOA**

Sandra Feijóo Ribeiro Videira e Castro

**DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE
MESTRE EM MICROBIOLOGIA MÉDICA**

JUNHO DE 2010



UNIVERSIDADE NOVA DE LISBOA

Instituto de Higiene e Medicina Tropical

Unidade de Ensino e Investigação de Virologia

**DIVERSIDADE GENÉTICA E RESISTÊNCIA AOS
ANTI-RETROVIRAIS INIBIDORES ENZIMÁTICOS DE
VÍRUS DA IMUNODEFICIÊNCIA HUMANA TIPO 1 (HIV-1) DE
UMA POPULAÇÃO DE TOXICODEPENDENTES POR
VIA ENDOVENOSA DA GRANDE LISBOA**

Sandra Feijóo Ribeiro Videira e Castro

**DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE
MESTRE EM MICROBIOLOGIA MÉDICA**

Orientador: Professor Doutor João Piedade

Junho de 2010

APRESENTAÇÕES CIENTÍFICAS EFECTUADAS NO ÂMBITO DESTE TRABALHO:

Sandra Videira e Castro, Teresa Venenno, Aida Esteves, Ricardo Parreira, Elizabeth Pádua, João Piedade. 2009. Genetic diversity and complexity of anti-retroviral drug resistance in HIV-1 strains circulating among intravenous drug users in the Greater Lisbon. Congresso MicroBiotec'09, Vilamoura, Portugal, 28-30 Novembro. (Anexo 1).

(Abstract number: 148)

S. Videira e Castro, T. Venenno, A. Esteves, R. Parreira, E. Pádua, J. Piedade. 2010. Genetic diversity and complexity of anti-retroviral drug resistance in HIV-1 strains circulating among intravenous drug users in the Greater Lisbon. 4th European Congress of Virology, Cernobbio, Como, Italy, April 7-11 (Anexo 1).

(Abstract number: 827)

AGRADECIMENTOS

O meu especial agradecimento vai para o meu Orientador Professor João Piedade, por ter acreditado em mim, pela sua paciência, amizade, disponibilidade incondicional e particularmente pela sua excelente orientação e cuidadosa revisão deste manuscrito;

À Professora Aida Esteves e Professor Ricardo Parreira pela simpatia, boa disposição, disponibilidade e pelo conhecimento partilhado durante os dois anos de mestrado;

À Teresinha, não só pela preciosa ajuda laboratorial, pelo carinho e amizade com que desde logo me acolheu;

À Fernanda pela sua prontidão, motivação e sorriso constante;

À Doutora Elizabeth Pádua e Ana Filipa Francisco do Instituto Nacional Dr. Ricardo Jorge e às enfermeiras Conceição Salavessa e Ania Balça do Centro de Atendimento a Toxicodependentes (CAT-Taipas), o meu profundo obrigado pelo fornecimento das amostras e dos respectivos dados epidemiológicos;

À Fundação *GlaxoSmithKline* das Ciências da Saúde por ter financiado o presente trabalho;

A todos os meus amigos pela constante motivação e por todos os momentos de convívio partilhados;

À minha família, em especial aos meus pais e ao meu irmão o meu particular obrigada, pelo imenso carinho e compreensão que sempre demonstraram.

Ao Jorge, pelo carinho e paciência que sempre demonstrou ao longo destes dois anos o meu especial obrigada por nunca ter desistido de mim e a quem eu quero dedicar esta tese. TQM!

Muito Obrigado a Todos!

RESUMO

Neste estudo procedeu-se à caracterização da diversidade genética das regiões codificantes da protease (PR), transcriptase reversa (RT) e integrase (IN) do gene *pol* do vírus da imunodeficiência humana tipo 1 (HIV-1), bem como à pesquisa de polimorfismos genéticos associados à diminuição da susceptibilidade aos anti-retrovirais inibidores enzimáticos, circulante numa população de 51 indivíduos utilizadores de drogas por via endovenosa (IDUs) da Grande Lisboa.

Em termos globais, a análise filogenética realizada com base em 38 sequências nucleotídicas concatenadas revelou que 12 (31,6%), 13 (34,2%) e 13 (34,2%) das sequências analisadas eram dos subtipos B, G/CRF14_BG e de formas genéticas não-B/não-G (1 F1, 4 CRF02_AG e 8 formas recombinantes únicas), respectivamente. Relativamente à pesquisa de mutações associadas a resistência (perfil genotípico), foram encontradas 15, presentes em 50,0% (22/44) dos indivíduos, com uma distribuição de 1-3/indivíduo (4, todas acessórias, na PR; 6 na RT; 5 na IN, uma principal e 4 acessórias). Todavia, apenas 26,7% (4/15) dessas mutações conferiam uma expressão fenotípica de resistência a uma das classes de inibidores (da RT ou IN), em 9,1% (4/44) dos indivíduos. Foi ainda observada uma elevada frequência de outros polimorfismos genéticos, mais frequentes em subtipos não-B, alguns dos quais considerados assinaturas de subtipo.

A homologia genética total ou parcial com os subtipos B e G, reconhecida neste estudo, reflectirá a sua predominância na população de IDUs. Este resultado sugere também que a epidemia de HIV-1 em Portugal poderá estar a evoluir para um padrão epidemiológico singular, no qual os subtipos B, G e suas formas recombinantes predominam.

A proporção observada de indivíduos portadores de vírus com mutações associadas a resistência, mesmo em indivíduos *naïve* relativamente à terapêutica, indica que a sua transmissão se encontra em curso entre a população de IDUs, tendo, certamente, um impacto relevante ao nível da abordagem terapêutica e monitorização da infecção.

ABSTRACT

In this study, we intended to evaluate the genetic diversity of protease (PR), reverse transcriptase (RT) and integrase (IN) coding regions of human immunodeficiency virus type 1 (HIV-1) *pol* gene, as well as, to describe the genetic polymorphisms associated with decreased susceptibility to enzyme inhibitor antiretrovirals, in strains circulating in a population of 51 intravenous drug users (IDUs) from the Greater Lisbon.

Overall, the phylogenetic analysis, based on 38 concatenated sequences, revealed that 12 (31.6%), 13 (34.2%) and 13 (34.2%) of the sequences analyzed were classified as subtypes B, G/CRF14_BG and other non-B/non-G genetic forms (1 F1, 4 CRF02_AG and 8 unique recombinant forms), respectively. The search for resistance-associated genetic polymorphisms (genotypic profile) revealed 15 mutations, present in 50.0% (22/44) of the subjects, with a distribution of 1-3/individual (4 minor in PR; 6 in RT; 5 in IN, one major and 4 minor). However, only 26.7% (4/15) of these mutations confer a phenotypic profile of resistance and just to one class of inhibitors (RT or IN), being found in 9.1% (4/44) of the subjects. It was also observed a high frequency of other genetic polymorphisms, most often in non-B subtypes, some of which are considered subtype signatures.

The total or partial genetic homology of our sequences with subtypes B and G, identified in this study, reflects their prevalence in this population of IDUs. This finding suggests that the HIV-1 epidemics in Portugal may be evolving into a unique epidemiological pattern, in which subtypes B, G and their recombinant forms predominate.

The observed ratio of subjects infected with viruses carrying mutations associated with resistance, even in therapy-naïve individuals, indicates that the transmission is ongoing among IDUs. This fact will surely have an impact on antiretroviral therapy implementation and monitoring of infection in this population.

ÍNDICE GERAL

APRESENTAÇÕES CIENTÍFICAS EFECTUADAS NO ÂMBITO DESTE TRABALHO.....	i
AGRADECIMENTOS.....	ii
RESUMO.....	iii
ABSTRACT.....	iv
ÍNDICE GERAL.....	v
ÍNDICE DE FIGURAS.....	viii
ÍNDICE DE TABELAS.....	xii
LISTA DE ABREVIATURAS.....	xiii
 1. INTRODUÇÃO	
1.1 Descrição da Síndrome da Imunodeficiência Adquirida (SIDA).....	2
1.2 Descoberta do Vírus da Imunodeficiência Humana (HIV).....	2
1.3 Origem Biológica do HIV.....	3
1.4 O Início da Pandemia.....	4
1.5 Taxonomia e Características Gerais.....	5
1.6 Aspectos Epidemiológicos.....	6
1.6.1 A nível mundial.....	6
1.6.2 A situação actual em Portugal.....	7
1.7 Vírus da Imunodeficiência Humana Tipo 1 (HIV-1): Organização Estrutural e Genómica.....	8
1.8 Ciclo Replicativo Viral.....	11
1.9 Diversidade Genética.....	16
1.9.1 Classificação do HIV-1.....	16
1.9.1.1 Formas recombinantes do HIV-1 do grupo M.....	17
1.9.2 Distribuição geográfica do HIV-1 do grupo M.....	18
1.10 Mecanismos Responsáveis pela Diversidade Genética.....	19
1.11 Proteínas Codificadas Pelo Gene <i>pol</i>	21
1.11.1. Protease.....	21
1.11.2 Transcriptase reversa.....	22
1.11.3 Integrase.....	24
1.12 Terapêutica Anti-retroviral.....	26
1.12.1 Inibidores da transcriptase reversa.....	27
1.12.1.1 Inibidores nucleosídicos.....	27
1.12.1.2 Inibidores não nucleosídicos.....	28
1.12.2 Inibidores da protease.....	29
1.12.3 Inibidor da integrase: raltegravir.....	30
1.12.4 Terapêutica anti-retroviral de alta eficácia (HAART).....	33

1.13	Resistência Associada a Terapêutica Anti-retroviral e Diversidade Genética.....	33
1.14	Monitorização de Resistências aos Anti-Retrovirais.....	35
1.15	Caracterização de Mutações Associadas à Resistência.....	36
1.15.1	Mutações associadas à resistência aos PIs.....	37
1.15.2	Mutações associadas à resistência aos NRTIs.....	38
1.15.3	Mutações associadas à resistência aos NNRTIs.....	40
1.15.4	Mutações associadas à resistência ao raltegravir.....	41
1.16	Objectivos.....	42
2. MATERIAL E MÉTODOS		
2.1	Caracterização Geral da População Estudada.....	44
2.1.1	Origem das amostras biológicas.....	44
2.2	Processamento das Amostras Biológicas.....	45
2.3	Pesquisa de Anticorpos Anti-HIV.....	45
2.4	Extracção do RNA Viral a Partir de Plasma.....	46
2.5	Origem e Preparação dos <i>Primers</i>	47
2.6	Retrotranscrição e Amplificação do Gene <i>pol</i>	48
2.6.1	Planificação do protocolo de amplificação por RT- <i>nested</i> PCR.....	49
2.6.1.1	Retrotranscrição e primeira reacção de amplificação.....	51
2.6.1.2	Segunda reacção de amplificação.....	52
2.6.1.3	Condições de retrotranscrição <i>in vitro</i> /amplificação por <i>nested</i> PCR.....	52
2.6.2	Retrotranscrição com <i>primers</i> hexaméricos aleatórios.....	53
2.6.2.1	Amplificação do cDNA por <i>nested</i> PCR.....	55
2.7	Observação dos Fragmentos de DNA Amplificados por RT- <i>Nested</i> PCR.....	55
2.8	Purificação dos Fragmentos de DNA Amplificados.....	55
2.9	Clonagem dos Produtos de PCR no Vector Plasmídico <i>pGEM[®]-T Easy</i>	57
2.9.1	Caracterização do vector plasmídico <i>pGEM[®]-T Easy</i>	57
2.9.2	Etapas do processo de clonagem.....	58
2.10	Extracção do DNA Plasmídico Pelo Método de Lise Alcalina.....	59
2.10.1	Observação das moléculas de DNA plasmídico.....	60
2.11	Hidrólise do DNA Plasmídico com a Enzima de Restrição <i>EcoRI</i>	60
2.12	Extracção e Purificação do DNA Plasmídico para Sequenciação.....	61
2.13	Sequenciação do DNA.....	61
2.14	Análise Bioinformática de Sequências do HIV-1.....	62
2.14.1	Sequências nucleotídicas.....	62
2.14.1.1	Análise dos cromatogramas e edição das sequências.....	62
2.14.1.2	Pesquisa de homologia genética.....	62
2.14.1.3	Determinação automática dos genótipos virais.....	62
2.14.1.4	Construção de árvores filogenéticas.....	63
2.14.1.5	Análise de sequências divergentes.....	63

2.14.1.6	Análise das distâncias genéticas.....	64
2.14.2	Sequências de aminoácidos.....	64
2.15	Análise Estatística.....	65
3.	RESULTADOS	
3.1	Caracterização Epidemiológica da População Estudada	67
3.2	Caracterização dos Segmentos de DNA a Amplificar.....	69
3.3	Amplificação por RT-Nested PCR.....	70
3.3.1	Retrotranscrição <i>in vitro</i> com <i>primers</i> específicos.....	70
3.3.2	Retrotranscrição <i>in vitro</i> com <i>primers</i> hexaméricos aleatórios.....	71
3.4	Clonagem no Vector Plasmídico <i>pGEM®-T Easy</i>	73
3.4.1	Seleção de plasmídeos recombinantes.....	75
3.5	Análise Bioinformática de Sequências Nucleotídicas.....	76
3.5.1	Análise dos cromatogramas e edição das sequências.....	76
3.5.2.	Pesquisa de homologia genética.....	77
3.5.3	Análise filogenética.....	78
3.5.3.1	Determinação automática de genótipos virais.....	78
3.5.3.2	Determinação de genótipos virais após construção de árvores filogenéticas	80
3.6	Análise Bioinformática de Sequências de Aminoácidos	93
3.6.1	Caracterização das mutações associadas a resistência.....	94
3.6.1.1	Região da protease.....	95
3.6.1.2	Região da transcriptase reversa.....	96
3.6.1.3	Região da integrase.....	98
3.6.2	Caracterização de outros polimorfismos genéticos.....	100
3.6.2.1	Região da protease.....	101
3.6.2.2	Região da transcriptase reversa.....	102
3.6.2.3	Região da integrase.....	103
4.	DISCUSSÃO E CONCLUSÕES.....	105
5.	REFERÊNCIAS BIBLIOGRÁFICAS.....	130
6.	ANEXOS.....	153

ÍNDICE DE FIGURAS

1. INTRODUÇÃO

Figura 1.1	Vias de transmissão plausível de ancestrais do HIV a partir de primatas infectados com SIV.....	3
Figura 1.2	Coincidência geográfica entre o habitat natural de <i>Pan troglodytes troglodytes</i> e de <i>Cercocebus atys</i> e, respectivamente, as áreas de infecção original da espécie humana por HIV-1 e HIV-2.....	4
Figura 1.3	Relação filogenética entre o HIV-1/SIVcpz e HIV-2/SIVsmm...	4
Figura 1.4	Distribuição do número de casos de infecção HIV/SIDA a nível global.....	6
Figura 1.5	Principais vias de transmissão do HIV em Portugal.....	7
Figura 1.6	Microfotografia electrónica e ilustração esquemática de um virião de HIV-1...	8
Figura 1.7	Representação esquemática da organização genómica do HIV-1.....	9
Figura 1.8	Modelo esquemático das fases iniciais da entrada do HIV-1 na célula.....	11
Figura 1.9	Representação esquemática do complexo ribonucleoproteico RTC.....	12
Figura 1.10	Entrada do PIC para o interior do núcleo.....	13
Figura 1.11	Representação esquemática das três classes de mRNAs virais sintetizados durante o ciclo replicativo, baseado no processamento de <i>splicing</i> sofrido.....	14
Figura 1.12	Diferentes formas do HIV-1: forma Imatura e forma Madura.....	15
Figura 1.13	Representação esquemática do ciclo replicativo do HIV-1.....	16
Figura 1.14	Frequência das diferentes formas genéticas das sequências de HIV-1 com origem em Portugal.....	19
Figura 1.15	Ilustração do modelo estrutural da protease do HIV-1.....	21
Figura 1.16	Representação da estrutura do heterodímero da RT do HIV-1.....	22
Figura 1.17	Ilustração do modelo estrutural da integrase do HIV-1.....	24
Figura 1.18	Ilustração esquemática do mecanismo de acção dos NRTIs.....	27
Figura 1.19	Representação esquemática do mecanismo de acção dos NNRTIs.....	28
Figura 1.20	Representação esquemática de quatro alvos terapêuticos plausíveis para os inibidores da transcriptase reversa.....	29
Figura 1.21	Ilustração esquemática do mecanismo de acção dos inibidores da protease.....	29
Figura 1.22	Representação esquemática dos quatro potenciais alvos terapêuticos dos inibidores da protease.....	30
Figura 1.23	Representação esquemática do mecanismo de acção do raltegravir.....	31
Figura 1.24	Representação esquemática de quatro alvos terapêuticos possíveis para os inibidores da integrase.....	32
Figura 1.25	Listagem compilada pela IAS-USA das mutações primárias e secundárias na região codificante da protease do gene <i>pol</i> associadas à resistência aos PIs.....	38
Figura 1.26	Listagem compilada pela IAS-USA das mutações na região codificante da	

	transcritase reversa do gene <i>pol</i> associadas à resistência aos NRTIs.....	39
Figura 1.27	Listagem compilada pela IAS-USA das mutações na região codificante da transcritase reversa do gene <i>pol</i> associadas à resistência aos NNRTIs.....	40
Figura 1.28	Listagem compilada pela IAS-USA das mutações primárias na região codificante da integrase do gene <i>pol</i> associadas a resistência ao raltegravir.....	41
2. MATERIAL E MÉTODOS		
Figura 2.1	Representação esquemática do princípio do sistema <i>Titan One Tube RT-PCR</i> ...	49
Figura 2.2	Representação esquemática da amplificação das regiões da protease, transcritase reversa e integrase do gene <i>pol</i> do HIV-1.....	50
Figura 2.3	Mapa físico do vector plasmídico de 3015 pb, <i>pGEM-T® Easy</i>	57
3. RESULTADOS		
Figura 3.1	Representação gráfica dos resultados da análise de distribuição relativa das diferentes variáveis.....	67
Figura 3.2	Representação gráfica dos resultados obtidos por RT-nested PCR (retrotranscrição <i>in vitro</i> com <i>primers</i> específicos) para as três regiões do gene <i>pol</i>	70
Figura 3.3	Visualização dos produtos de amplificação por RT-nested PCR das regiões da protease e transcritase reversa (à esquerda) e da integrase (à direita) do gene <i>pol</i> , após electroforese em gel de agarose a 1,2% (p/v), em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml).....	71
Figura 3.4	Visualização dos produtos de amplificação por RT-nested PCR das regiões da protease, transcritase reversa e integrase do gene <i>pol</i> , após electroforese em gel de agarose a 1,2 % (p/v) em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml).....	72
Figura 3.5	Visualização dos produtos de amplificação por RT-nested PCR das regiões da protease e integrase do gene <i>pol</i> (amostra PT_TP85), após electroforese em gel de agarose a 1,2% (p/v), em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml).....	72
Figura 3.6	Representação gráfica dos resultados obtidos por RT-nested PCR (retrotranscrição <i>in vitro</i> com <i>primers</i> específicos ou hexaméricos aleatórios) para as três regiões do gene <i>pol</i>	73
Figura 3.7	Rastreio de clones bacterianos recombinantes baseado no sistema de α -complementação branco/azul, após ligação dos produtos de PCR no vector plasmídico <i>pGEM®-T Easy</i> , transformação por choque térmico e plaqueamento em meio LB agar com ampicilina, X-gal e IPTG. O aparecimento de colónias brancas sugere a inserção do fragmento em estudo no local de clonagem múltipla no gene <i>lacZ</i>	74

Figura 3.8	Visualização dos DNAs plasmídicos (à esquerda) e dos seus perfis de digestão enzimática com EcoRI (à direita), após electroforese em gel de agarose a 1,0-1,2% (p/v), em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml). Os DNAs plasmídicos recombinantes putativos seleccionados para hidrólise enzimática encontram-se assinalados com setas de cor azul (à esquerda).....	75
Figura 3.9	A) Relações filogenéticas entre 94 sequências nucleotídicas correspondentes à região da protease do gene <i>pol</i> do HIV-1, a partir de um alinhamento de 297 nucleótidos. B) À direita da árvore filogenética encontram-se representados a análise de <i>bootscanning</i>.....	82
Figura 3.10	A) Relações filogenéticas entre 93 sequências nucleotídicas correspondentes à região da transcriptase reversa do gene <i>pol</i> do HIV-1, a partir de um alinhamento de 598 nucleótidos. B) À direita da árvore filogenética encontram-se representados a análise de <i>bootscanning</i>.....	83
Figura 3.11	A) Relações filogenéticas entre 80 sequências nucleotídicas correspondentes à região da integrase do gene <i>pol</i> do HIV-1, a partir de um alinhamento de 813 nucleótidos. B) À direita da árvore filogenética encontram-se representados a análise de <i>bootscanning</i>.....	84
Figura 3.12	Identificação das sequências que apresentavam posicionamentos ambíguos na respectiva árvore filogenética dificultando a sua genotipagem.....	86
Figura 3.13	A) Relações filogenéticas entre 76 sequências nucleotídicas correspondentes às sequências concatenadas (PR+RT+IN), do gene <i>pol</i> do HIV-1, a partir de um alinhamento de 1707 nucleótidos. B) À direita da árvore filogenética encontram-se representados a análise de <i>bootscanning</i>.....	88
Figura 3.14	Representação gráfica da proporção relativa dos subtipos genéticos do HIV-1 obtidos a partir da análise filogenética de 38 sequências nucleotídicas concatenadas (PR+RT+IN).....	93
Figura 3.15	Representação gráfica dos resultados (em percentagem) da análise de distribuição relativa à terapêutica anti-retroviral para os 44 indivíduos infectados com pelo menos uma região amplificada do genoma do HIV-1.....	93
Figura 3.16	Representação esquemática das mutações de resistência acessórias presentes na região da transcriptase reversa do gene <i>pol</i> no grupo de 41 sequências em estudo.....	95
Figura 3.17	Representação gráfica dos perfis fenotípicos associados às mutações de resistência aos inibidores da protease encontradas em sequências nucleotídicas dos subtipos B e não-B.....	96
Figura 3.18	Representação esquemática das mutações de resistência aos inibidores não nucleosídicos da transcriptase reversa presentes na região da transcriptase reversa do gene <i>pol</i> no grupo de 41 sequências em estudo.....	96

Figura 3.19	Representação gráfica dos perfis fenotípicos associados às mutações de resistência aos inibidores da transcriptase reversa (NNRTIs) encontradas em sequências nucleotídicas dos subtipos B e não-B.....	98
Figura 3.20	Representação esquemática das mutações de resistência primárias e acessórias presentes na região da integrase do gene <i>pol</i>, no grupo de 41 sequências em estudo.....	98
Figura 3.21	Representação gráfica dos perfis fenotípicos associados às mutações de resistência aos inibidores da integrase encontradas em sequências nucleotídicas dos subtipos B e não-B.....	99
Figura 3.22	Representação gráfica da distribuição do número de mutações associadas a resistência, encontradas nas sequências da protease, transcriptase reversas e integrase do gene <i>pol</i>.....	100
Figura 3.23	Representação esquemática dos polimorfismos genéticos não associados a resistência aos anti-retrovirais da região da protease do gene <i>pol</i>, associados, de um modo significativo, aos subtipos B (1) e não-B (2), presentes no grupo de 41 sequências em estudo.....	102
Figura 3.24	Representação esquemática dos polimorfismos genéticos não associados a resistência aos anti-retrovirais da região da transcriptase reversa do gene <i>pol</i>, associados, de um modo significativo, aos subtipos B (1) e não-B (2), presentes no grupo de 41 sequências em estudo.....	103
Figura 3.25	Representação esquemática dos polimorfismos genéticos não associados a resistência aos anti-retrovirais da região da integrase do gene <i>pol</i>, associados, de um modo significativo, aos subtipos B (1) e não-B (2), presentes no grupo de 41 sequências em estudo.....	104

ÍNDICE DE TABELAS

2. MATERIAL E MÉTODOS

Tabela 2.1	<i>Primers</i> utilizados para a 1ª e 2ª reacções de RT- <i>nested</i> PCR.....	48
Tabela 2.2	Condições de retrotranscrição <i>in vitro</i> e de amplificação (1ª reacção de PCR) do RNA viral, para as regiões da protease, transcriptase reversa e integrase do gene <i>pol</i> do HIV-1.....	53
Tabela 2.3	Condições de amplificação (2ª reacção de PCR) do cDNA matriz, para as regiões da protease, transcriptase reversa e integrase do gene <i>pol</i> do HIV-1.....	53
Tabela 2.4	Condições de retrotranscrição <i>in vitro</i> do RNA viral com <i>primers</i> hexaméricos aleatórios.....	54

3. RESULTADOS

Tabela 3.1	Caracterização breve dos fragmentos do gene <i>pol</i> do HIV-1 a amplificar por RT- <i>nested</i> PCR.....	69
Tabela 3.2	Número total de sequências nucleotídicas editadas e não editadas.....	77
Tabela 3.3	Tamanho do fragmento amplificado por <i>nested</i> PCR, tamanho das sequências editadas e respectiva localização em HXB2 e tamanho da região estudada relativamente a HXB2.....	77
Tabela 3.4	Resultados da determinação automática dos genótipos virais [<i>REGA HIV-1 & 2 Automated Subtyping Tool</i> , v.2.0].....	79
Tabela 3.5	Número de sequências nucleotídicas utilizadas na construção das árvores filogenéticas.....	81
Tabela 3.6	Resultados de genotipagem após construção de árvores filogenéticas.....	85
Tabela 3.7	Resultados de genotipagem após construção da árvore filogenética com as sequências concatenadas (PR+RT+IN).....	89
Tabela 3.8	Frequências absolutas das mutações de resistência aos inibidores da protease por subtipo (B <i>vs.</i> não-B), com indicação do valor de <i>p</i> associado à distribuição apresentada.....	95
Tabela 3.9	Frequências absolutas das mutações de resistência aos inibidores não nucleosídicos da transcriptase reversa por subtipo (B <i>vs.</i> não-B), com indicação do valor de <i>p</i> associado à distribuição apresentada.....	97
Tabela 3.10	Frequências absolutas das mutações de resistência aos inibidores da integrase por subtipo (B <i>vs.</i> não-B), com indicação do valor de <i>p</i> associado à distribuição apresentada.....	99
Tabela 3.11	Informação relevante sobre as sequências de HIV-1 que possuíam pelo menos uma mutação associada a um perfil fenotípico de resistência nas regiões do genoma em estudo (PR, RT e IN).....	100

LISTA DE ABREVIATURAS

A	Adenina (base orgânica constituinte dos nucleótidos)
AMV	Vírus da mieloblastose aviária, do inglês <i>avian myeloblastosis virus</i>
ARV	Retrovírus associado à SIDA, do inglês <i>AIDS-associated retrovirus</i>
ATP	Adenosina 5' trifosfato do inglês, <i>adenosine-5'-triphosphate</i>
bla	Gene codificante das β -lactamases, do inglês <i>beta-lactamase</i>
BLAST	Ferramenta bioinformática, do inglês <i>Basic Local Alignment Search Tool</i>
BSA	Albumina de soro bovino, do inglês <i>bovine serum albumin</i>
C	Citosina (base orgânica constituinte dos nucleótidos)
CA	Proteína da cápside do HIV-1 (ou p24)
CAT-TAIPAS (IDT)	Centro de Atendimento de Toxicodependentes das Taipas, Instituto da Droga e da Toxicodependência
CCR5	Receptor de β -quimiocinas da classe CC, do inglês <i>CC chemokine receptor 5</i>
CDC	Centro de Controlo e Prevenção de Doenças, do inglês <i>Centers for Disease Control and Prevention</i>
CD4	Receptor celular, do inglês <i>cluster of differentiation type 4</i>
CRF	Forma recombinante circulante, do inglês <i>circulating recombinant form</i>
CV	Carga viral
CXCR4	Receptor de β -quimiocinas da classe CXC, do inglês <i>CXC chemokine receptor 4</i>
DC	Domínio catalítico da integrase
DC-SIGN	Molécula não integrina, captadora da molécula de adesão intercelular 3 e específica das células dendríticas, do inglês <i>dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin</i>
DCT	Domínio C-terminal da integrase
DDE	Tríade catalítica de resíduos acídicos da integrase no seu centro activo, de aspartato (D)-aspartato (D)-glutamato (E)
DEPC	Dietilpirocarbonato
DMSO	Dimetilsulfóxido
DNA	Ácido desoxirribonucleico, do inglês <i>deoxyribonucleic acid</i>
DNT	Domínio N-terminal da integrase
dNTPs	Desoxirribonucleótidos 5'-trifosfatados
D.O.₆₀₀	Densidade óptica medida a 600 nm de comprimento de onda
DTT	Ditiotreitol
EDTA	Ácido etilenodiaminotetracético, do inglês <i>ethylene diamine tetracetic acid</i>
env	Gene codificante das glicoproteínas do invólucro do HIV, do inglês <i>envelope</i>
FDA	Entidade norte-americana reguladora alimentar e do medicamento, do inglês

Food and Drug Administration

G	Guanina (base orgânica constituinte dos nucleótidos)
gag	Gene codificante das proteínas da estrutura interna do HIV, do inglês <i>group specific antigen</i>
gp	Glicoproteína
Grupos M/N/O	Classificação adoptada para o HIV-1, do inglês <i>main/major, new/non-M, non-O e outlier</i> , respectivamente
HAART	Terapêutica anti-retroviral de alta eficácia, do inglês <i>highly active antiretroviral therapy</i>
HFF (AS)	Hospital Professor Doutor Fernando da Fonseca (Amadora-Sintra)
HHCC	Motivo de histidina (H)-histidina (H)-cisteína (C)-cisteína (C) da integrase
HIV	Vírus da imunodeficiência humana, do inglês <i>human immunodeficiency virus</i>
HMA	Ensaio de mobilidade de heteroduplexes, do inglês <i>heteroduplex mobility assay</i>
HS	De elevada salinidade, do inglês <i>high salt</i>
HTLV III	Vírus T-linfotrópico humano tipo III, do inglês <i>human T-lymphotropic virus type III</i>
IC	Valor de concentração mínima inibitória, do inglês <i>inhibitory concentration</i>
IDUs	Utilizadores de drogas por via endovenosa, do inglês <i>intravenous drug users</i>
IN	Integrase do HIV-1 (ou p32)
INIs	Inibidores da integrase, do inglês <i>integrase inhibitors</i>
INSA	Instituto Nacional de Saúde Doutor Ricardo Jorge
IPTG	Isopropil-β-D-tiogalactopiranosídeo
LAV	Vírus associado a linfadenopatias, do inglês <i>lymphadenopathy-associated virus</i>
LB	Meio de lisogenia, do inglês <i>lysogeny broth</i> (frequentemente designado de Luria-Bertani)
LEDGF/p75	Co-factor celular do processo de integração viral, do inglês <i>lens epithelium-derived growth factor</i>
LS	De baixa salinidade, do inglês <i>low salt</i>
LTR	Regiões repetitivas terminais longas do genoma proviral do HIV, do inglês <i>long terminal repeats</i>
M	Molaridade (número de moles por litro de solução)
MA	Proteína da matriz do HIV-1 (ou p17)
MDBD	Região inespecífica de ligação ao DNA, do inglês <i>minimal nonspecific DNA binding region</i>
MEGA	Ferramenta bioinformática, do inglês <i>Molecular Evolutionary Genetics Analysis</i>

ml	Mililitro
mM	Milimolar
mRNA	Ácido ribonucleico mensageiro, do inglês <i>messenger ribonucleic acid</i>
N	Normalidade (número de equivalentes de $[H^+]$ por litro de solução)
NC	Proteína da nucleocápside do HIV-1 (ou p7)
nef	Gene acessório do HIV, do inglês <i>negative factor</i>
NF-kB	Factor celular de transcrição nuclear, do inglês <i>nuclear factor kappa B</i>
NNRTIs	Inibidores não nucleosídicos da transcriptase reversa, do inglês <i>non-nucleoside reverse transcriptase inhibitors</i>
NRTIs	Inibidores nucleosídicos da transcriptase reversa, do inglês <i>nucleoside reverse transcriptase inhibitors</i>
OMS	Organização Mundial da Saúde
pb	Pares de bases
PBMC	Células mononucleadas do sangue periférico, do inglês <i>peripheral blood mononuclear cells</i>
PCR	Reacção em cadeia da polimersase, do inglês <i>polymerase chain reaction</i>
PEG	Polietilenoglicol
pH	Simétrico do logaritmo decimal da concentração hidrogeniônica de uma solução
PIs	Inibidores da protease, do inglês <i>protease inhibitors</i>
PIC	Complexo de pré-integração, do inglês <i>pre-integration complex</i>
pmol	Picomole = 10^{-12} moles
pol	Gene codificante das proteínas de natureza enzimática do HIV, do inglês <i>polymerase</i>
PPT	Trecho de polipurinas, do inglês <i>polypurine tract</i>
PR	Protease do HIV-1 (ou p11)
p/v	Peso por volume
R	Região não codificante repetida situada nas extremidades do genoma de HIV-1, do inglês <i>repeated sequence</i>
rev	Gene regulador do HIV, do inglês <i>regulator of expression of viral proteins</i>
RNA	Ácido ribonucleico, do inglês <i>ribonucleic acid</i>
RNase	Ribonuclease
rpm	Revoluções por minuto
RRE	Elemento de ligação da proteína viral Rev, do inglês <i>Rev responsive element</i>
RT	Transcritase reversa do HIV-1 (ou p66/p51)
RT-nested PCR	Retrotranscrição e amplificação por <i>nested PCR</i> , do inglês <i>reverse transcription-nested polymerase chain reaction</i>
RTC	Complexo ribonucleoproteico de transcrição reversa, do inglês <i>reverse</i>

	<i>transcription complex</i>
SDS	Dodecil sulfato de sódio, do inglês <i>sodium dodecyl sulphate</i>
SIDA	Síndrome da imunodeficiência adquirida
SIV	Vírus da imunodeficiência símia, do inglês <i>simian immunodeficiency virus</i>
SU	Glicoproteína de superfície do invólucro do HIV-1 (ou gp120)
SUDI	Ferramenta bioinformática, do inglês <i>Subtyping Distance Tool</i>
T	Timina (base orgânica constituinte dos nucleótidos)
TAE	Tampão Tris-acetato-EDTA
TAMs	Mutações associadas aos análogos de timidina, do inglês <i>thymidine analogue associated mutations</i>
TAR	Elemento de resposta por transativação por Tat, do inglês <i>transactivation response element</i>
<i>tat</i>	Gene regulador do HIV, do inglês <i>transactivator of gene expression</i>
TEG	Tampão Tris-EDTA-glicose
T_H	Temperatura de hibridação, do inglês <i>hybridization temperature</i>
T_M	Temperatura de fusão, do inglês <i>melting temperature</i>
TM	Glicoproteína transmembranar do invólucro do HIV-1 (ou gp41)
tRNA	Ácido ribonucleico de transferência, do inglês <i>transfer ribonucleic acid</i>
TSS	Solução de transformação e armazenamento, do inglês <i>transformation and storage solution</i>
U	Unidade de actividade enzimática
U	Região do genoma viral de genótipo desconhecido, do inglês <i>unknown</i>
U	Uracilo (base orgânica constituinte dos nucleótidos)
UNAIDS	Programa das Nações Unidas de Luta Contra a SIDA
URF	Forma recombinante única, do inglês <i>unique recombinant form</i>
U3	Região situada na extremidade 3' do genoma de HIV-1, do inglês <i>unique sequence 3</i>
U5	Região não codificante situada próximo da extremidade 5' do genoma de HIV-1, do inglês <i>unique sequence 5</i>
V	Volt
<i>vif</i>	Gene acessório do HIV, do inglês <i>viral infectivity factor</i>
<i>vpr</i>	Gene acessório do HIV, do inglês <i>viral protein R</i>
<i>vpu</i>	Gene acessório do HIV-1, do inglês <i>viral protein U</i>
v/v	Volume por volume
X-gal	5-bromo-4-cloro-3-indolil-β-D-tiogalactopiranósido
λ	Comprimento de onda
μ	Micro (prefixo que designa 10 ⁻⁶)
μg	Microgramas

μl	Microlitros
μM	Micromolar
°C	Grau Celsius

1.Introdução

1.1 Descrição da Síndrome da Imunodeficiência Adquirida (SIDA)

A partir de Outubro de 1980, num espaço de um ano, foram reportados ao CDC (do inglês, *Centers for Disease Control and Prevention*) de Atlanta, nos Estados Unidos da América (EUA), vários casos de doenças raras (por exemplo, candidíase das mucosas por *Candida albicans*, pneumonia por *Pneumocystis jirovecii*, infecções por citomegalovírus associadas ao sarcoma de Kaposi e linfadenopatias persistentes generalizadas), num grupo de indivíduos jovens, previamente saudáveis, que tinham em comum a sua homossexualidade (27, 28, 29, 57, 84). A elevada incidência de novos casos em diferentes grupos de risco criou condições para a identificação e caracterização de uma nova entidade clínica, que o CDC acabou por denominar síndrome da imunodeficiência adquirida (SIDA) (30). Esta síndrome tinha por base uma marcada disfunção progressiva do sistema imunitário, principalmente mediante a depleção dos linfócitos T auxiliares CD4+ (do inglês, *cluster of differentiation type 4*), encontrando-se geralmente associada ao aparecimento de infecções oportunistas e/ou neoplasias (200).

1.2 Descoberta do Vírus da Imunodeficiência Humana (HIV)

No ano de 1983, a equipa de Luc Montagnier, no Instituto Pasteur de Paris, isolou pela primeira vez o agente etiológico da SIDA, a partir de um gânglio linfático de um indivíduo homossexual com linfadenopatias generalizadas (7). Um pouco mais tarde, dois outros grupos de investigadores, liderados por Robert Gallo e Jay Levy, anunciam também o isolamento do vírus causador da SIDA, designando-o respectivamente por HTLV-III (do inglês, *human T-cell leukemia virus type 3*) e ARV (do inglês, *AIDS-associated retrovirus*), em oposição à designação inicial de LAV (do inglês, *lymphadenopathy-associated virus*) (38, 75, 117, 138, 168, 189, 190). Em Maio de 1986, foi proposta pelo Comité Internacional da Taxonomia dos Vírus uma nova designação, vírus da imunodeficiência humana, HIV (do inglês, *human immunodeficiency virus*) (40, 41). Nesse mesmo ano, Clavel *et al.* identificaram e isolaram, em indivíduos com SIDA, um vírus, geneticamente distinto do HIV então conhecido, prevalente em certas regiões da África Ocidental (38). Este novo vírus, embora um lentivírus, apresentava algumas diferenças nas suas características antigénicas e relações filogenéticas relativamente ao agente viral previamente identificado. Por esse motivo, o HIV passou a ser classificado em dois tipos: HIV-1 e

HIV-2 (37, 39). Embora partilhem características morfológicas e de organização genómica, o HIV-1 e o HIV-2 diferem entre si cerca de 60% ao nível da sequência nucleotídica (174). Enquanto o HIV-1 é responsável pela pandemia global, o HIV-2 encontra-se praticamente confinado a alguns países da África Ocidental, onde é endémico (187). Este último é ainda responsável por um número reduzido de casos de infecção na Europa e noutros continentes que ocorrem, sobretudo, em indivíduos oriundos da África Ocidental ou com ligações a esta região. O HIV-2 é menos patogénico e transmissível que o HIV-1, podendo isto estar na origem da sua menor prevalência e distribuição geográfica limitada (4, 129).

1.3 Origem Biológica do HIV

A história evolutiva do HIV tem sido reconstruída com grande detalhe através de estudos de inferência filogenética realizados com os vírus da imunodeficiência simia, SIV (do inglês, *simian immunodeficiency virus*), com os quais o HIV partilha homologia considerável (76, 86). Os dois tipos de vírus associados à SIDA foram introduzidos na espécie humana por eventos de transmissão de carácter zoonótico, sendo esta teoria sustentada por fortes evidências genéticas, filogenéticas e epidemiológicas (86, 196). A hipótese mais provável sugere que o SIV tenha sido transmitido ao Homem como resultado da

exposição directa ao sangue de primatas africanos naturalmente infectados (**Figura 1.1**). Este facto pode ter ocorrido aquando do manuseamento de carne de símios usada na alimentação ou pela sua domesticação como animais de estimação, que são práticas comuns na África Equatorial (34, 35). Existe uma elevada semelhança na organização estrutural do genoma do HIV-1 e do HIV-2 com os SIV de



Figura 1.1. Vias de transmissão plausível de ancestrais do HIV a partir de primatas infectados com SIV (86).

chimpanzés da subespécie *Pan troglodytes troglodytes* (SIVcpz_{P.t.t}) e dos mangabeis fuliginosos da espécie *Cercocebus atys* (SIVsmm), respectivamente (92, 99). Há, por outro lado, uma coincidência geográfica assinalável (**Figura 1.2**) entre o habitat natural dos primatas hospedeiros dos SIV filogeneticamente mais próximos dos HIV e as áreas de infecção original da espécie humana por estes últimos. O habitat natural de *Pan*

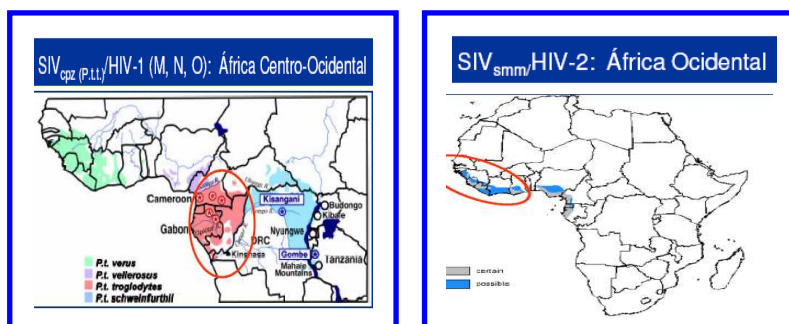


Figura 1.2. Coincidência geográfica entre o habitat natural de *Pan troglodytes troglodytes* (186) e de *Cercopithecus atys* (100) e, respectivamente, as áreas de infecção original da espécie humana por HIV-1 e HIV-2.

troglodytes troglodytes encontra-se na África Ocidental (Gabão, Guiné Equatorial, Camarões e República do Congo) coincidindo com as áreas de infecção ancestral do HIV-1. (76, 86).

Relativamente ao HIV-2, as áreas endêmicas estão localizadas na costa oeste africana (Senegal, Guiné-Bissau, Costa do Marfim e partes da Serra Leoa e da Libéria), o que coincide com o habitat natural da espécie *Cercopithecus atys* (35). Outra evidência relevante é a elevada prevalência de SIVcpz e SIVmm nos seus hospedeiros naturais, sem causar doença (35, 106, 158, 159), o que parece provar uma circulação de longa data dos vírus nessas espécies.

Finalmente, estudos filogenéticos comprovam que os dois tipos de HIV partilham uma maior proximidade evolutiva com os respectivos SIV, do que entre si (Figura 1.3), não sendo possível separá-los, nas árvores filogenéticas, de acordo com a respectiva espécie hospedeira de origem (35, 42, 76). Foi proposto que a infecção por HIV-1 tenha tido origem na transmissão cruzada de SIVcpz_{P.t.t.} em, pelo menos, três ocasiões diferentes (86), enquanto a infecção por HIV-2 terá tido origem na transmissão cruzada de SIVsmm, em, pelo menos, oito eventos independentes (92).

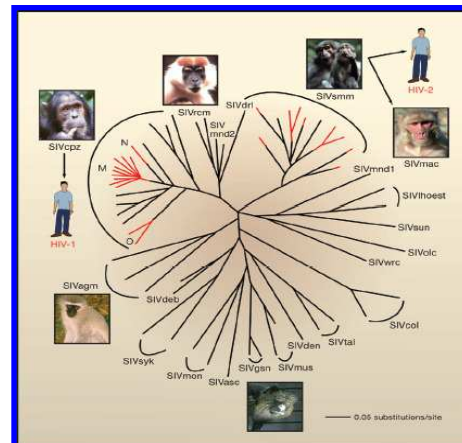


Figura 1.3. Relação filogenética entre HIV-1/SIVcpz e HIV-2/SIVsmm (93).

1.4 O Início da Pandemia

Dados actuais sugerem que a epidemia de HIV-1 terá tido o seu início na região ocidental da África Equatorial (206). A disponibilidade de um número crescente de sequências de estirpes virais com origem nessa área geográfica tem permitido o

desenvolvimento de modelos de evolução molecular, os quais procuram relacionar vários acontecimentos numa perspectiva temporal retrospectiva. Os cálculos iniciais apontavam para que a grande maioria das estirpes HIV-1 actuais tenha compartilhado um ancestral símio comum, o qual foi introduzido na população humana africana numa data próxima de 1930 e que a partir daí se tenha diversificado (112, 131 156, 246). Num estudo posterior, foi proposto que o HIV-1 já circulava na população humana antes dessa data, tendo-se estimado que o primeiro episódio de transmissão de carácter zoonótico terá ocorrido entre 1884 e 1924 (241).

1.5 Taxonomia e Características Gerais

Os vírus da imunodeficiência humana dos tipos 1 e 2 pertencem à família *Retroviridae*, subfamília *Orthoretrovirinae*, género *Lentivirus* e grupo dos Lentivírus dos Primatas. O genoma viral é de RNA (do inglês, *ribonucleic acid*) e contém 9 genes. Uma característica importante destes vírus é o passo de transcrição reversa no seu ciclo replicativo, dependente da actividade da transcriptase reversa. Esta enzima tem a capacidade de converter cada uma das moléculas de RNA genómico de cadeia simples numa molécula de DNA (do inglês, *deoxyribonucleic acid*), de cadeia dupla. Outra característica importante destes vírus é a capacidade de integração do seu genoma, na forma de DNA complementar, no genoma da célula hospedeira (DNA proviral), independentemente da fase do ciclo celular em que esta se encontre (24, 68). Ambos os tipos do HIV conduzem os indivíduos infectados ao estado de SIDA (145, 147), embora estudos clínicos longitudinais tenham demonstrado que a velocidade de progressão da infecção para estádios mais avançados de imunodepressão seja significativamente mais baixa no caso da infecção pelo HIV-2 (129).

A entrada do HIV no organismo, em especial se a infecção ocorrer ao nível das mucosas, depende essencialmente da sua interacção com células do sistema imunitário, nomeadamente, as células dendríticas [a ligação é feita através de uma molécula presente à superfície das células dendríticas, denominada *DC-SIGN* (do inglês, *dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin*)] (146). Mais tarde, durante a infecção, o HIV apresenta tropismo particular para outras células do sistema imunitário, principalmente para os linfócitos T CD4+ auxiliares e as células da linhagem monocitária. A sua especificidade deve-se essencialmente à utilização de receptores de membrana CD4, presentes nessas células, juntamente com receptores de

β -quimiocinas, maioritariamente CXCR4 ou CCR5, para os quais o vírus possui tropismo diferencial (70).

Nos últimos tempos, grandes avanços na descoberta de factores de restrição celulares envolvidos na protecção da infecção contra o HIV (imunidade inata) têm sido realizados. A proteína celular, APOBEC3G, por exemplo, converte resíduos de desoxicitidina em desoxiuridina nas moléculas nascentes de DNA de polaridade negativa, durante o mecanismo de retrotranscrição do RNA viral. Este factor induz a fixação de mutações ao nível do genoma proviral, fenómeno conhecido por hipermutação G→A, o que poderá revelar-se letal para o vírus (149, 181)

1.6 Aspectos Epidemiológicos

1.6.1 A nível mundial

De acordo com as estimativas da Organização Mundial da Saúde (OMS), o número total de pessoas a viver com a infecção HIV/SIDA, em Dezembro de 2008, era de 33,4 milhões [31,1-35,8 milhões]. Foi estimado que nesse ano ocorreram cerca de 2,7 milhões [2,4-3,0 milhões] de novas infecções e que 2,0 milhões [1,7-2,4 milhões] de indivíduos morreram por causas relacionadas com a SIDA. A população de adultos e crianças infectadas com o HIV encontra-se concentrada fundamentalmente no continente africano, principalmente na região subsariana (**Figura 1.4**). Esta região continua a ser a mais afectada, com 22,4 milhões [20,8-24,1 milhões] de indivíduos infectados e uma prevalência estimada de infecção na população adulta de 5,2% (225).



Figura 1.4. Distribuição do número de casos de infecção HIV/SIDA a nível global.

No que respeita às vias de transmissão do HIV, a via sexual continua a ser a mais importante, representando aproximadamente 85% dos casos de infecção a nível mundial (70, 212). Inicialmente, a transmissão do vírus encontrava-se especialmente associada às relações do tipo homossexual. Com o evoluir da epidemia, a frequência de transmissão heterossexual aumentou, tornando-se presentemente superior à do primeiro tipo (225).

A epidemia do HIV entre utilizadores de drogas por via endovenosa (IDUs, do inglês *intravenous drug users*) continua a grassar em diferentes locais da Europa

Oriental, América e Ásia (132, 238). A nível mundial, foi estimado que 16 milhões de indivíduos consomem drogas por via endovenosa e, destes, cerca de 3 milhões encontram-se infectados com HIV (132).

1.6.2 A situação actual em Portugal

A infecção de HIV/SIDA em Portugal possui características específicas que a distinguem da maior parte dos países europeus. De acordo com a classificação adoptada pela OMS, a epidemia portuguesa é do tipo concentrado. A prevalência estimada da

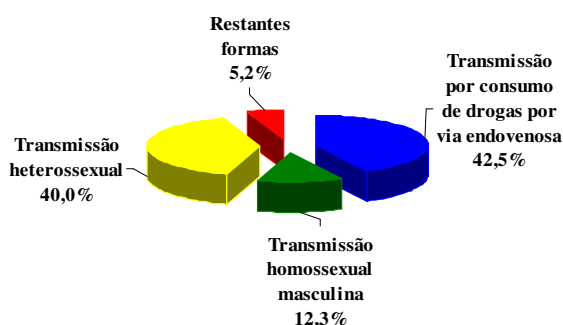


Figura 1.5. Principais vias de transmissão do HIV em Portugal (a 31 de Dezembro de 2008) (31).

infecção HIV/SIDA na população adulta portuguesa é de 0,5% [0,3-0,9%] (225), mas superior a 5% em pelo menos dois grupos vulneráveis, IDUs e reclusos (44). A 31 de Dezembro de 2008, encontravam-se notificados 34 888 casos de infecção HIV/SIDA em Portugal, nos

diferentes estádios de infecção (“casos acumulados”). Os dados epidemiológicos disponíveis indicam que o maior número de casos notificados corresponde a indivíduos infectados através do consumo de drogas por via endovenosa, abrangendo 42,5% de todas as notificações (**Figura 1.5**). O número de casos associados à infecção por transmissão sexual (heterossexual) representa 40,0% dos registos. A transmissão homossexual masculina representa 12,3% dos casos e as restantes formas de transmissão (hemofílicos, transfusionados, mãe-filho) correspondem a 5,2% do total em Portugal (31). O peso relativo das vias de transmissão da infecção tem-se modificado ao longo do tempo. Os IDUs representaram, desde o início da epidemia e até 1999, a maior proporção de infectados. No entanto, entre os casos notificados, tem-se vindo a observar um aumento da transmissão sexual (heterossexual) e uma diminuição da transmissão através do consumo de drogas por via endovenosa. Esta diminuição pode ser explicada pelo aumento relativo do número de casos de transmissão heterossexual, mas também pelo sucesso das estratégias implementadas para o controlo e prevenção da infecção por HIV/SIDA na população de IDUs (31).

Para que se possa prever e controlar eventuais modificações dos actuais padrões epidemiológicos da infecção HIV/SIDA em Portugal, é fundamental conhecer a

heterogeneidade molecular dos vírus presentes e estudar a sua correlação com os grupos de transmissão e origem geográfica da população.

1.7 Vírus da Imunodeficiência Humana Tipo 1 (HIV-1): Organização Estrutural e Genómica



As partículas virais maduras do HIV-1, quando observadas ao microscópio electrónico, apresentam uma estrutura esférica, de aproximadamente 100 nm de diâmetro e uma simetria complexa. Possuem um invólucro fosfolipídico, derivado da membrana citoplasmática das células hospedeiras, no qual se encontram inseridas duas glicoproteínas de origem viral,

derivadas de um precursor poliproteico (gp160). Após clivagem por proteases celulares, gp160 dá origem às glicoproteínas de superfície gp120/SU (do inglês, *surface protein*) e transmembranar gp41/TM (do inglês, *transmembrane protein*), as quais, no virião, se associam na forma de homotrímeros de heterodímeros. Estas são responsáveis pela mediação do processo de fusão do invólucro do vírus com a membrana citoplasmática, após a ligação a receptores específicos situados na superfície das células alvo. A superfície interna do invólucro encontra-se intimamente associada à matriz viral, constituída pela proteína p17/MA (do inglês, *matrix protein*), a qual é extremamente importante para a integridade estrutural do virião. A envolver o material genético do vírus encontra-se a cápside viral, na forma de cone truncado, constituída pela proteína p24/CA (do inglês, *capsid protein*). No interior desta, encontram-se o genoma, associado às proteínas da nucleocápside p7/NC (do inglês, *nucleocapsid protein*), as

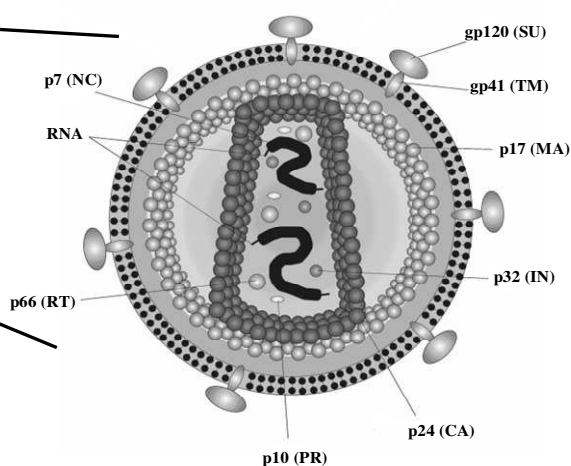


Figura 1.6. Microfotografia electrónica (lado esquerdo) e ilustração esquemática (lado direito) de um virião de HIV-1 (disponível em <http://www.stanford.edu/group/virus/retro/2005gongishmail/hiv1c.jpg> e http://www.pipelinedrugs.com/biotechnology_encyclopedia/250px-Hiv_gross.png, respectivamente).

proteínas de natureza enzimática transcriptase reversa p66/p51/RT (do inglês, *reverse transcriptase*), protease p11/PR (do inglês, *protease*) e integrase p32/IN (do inglês, *integrase*), e, por último, as proteínas virais acessórias Vif, Vpr e Nef (24, 199) (**Figura 1.6**).

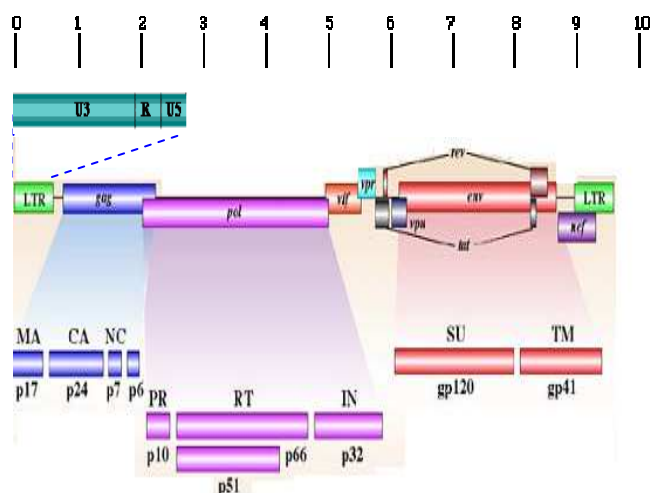


Figura 1.7. Representação esquemática da organização genômica do HIV-1 (disponível em www.stanford.edu/.../2005gongishmail/HIV.html). Estão identificadas as nove regiões codificantes, as extremidades LTR e as principais proteínas codificadas. Encontra-se igualmente ilustradas as regiões que constituem a LTR-5' (U3, R e U5). A escala acima encontra-se em kb.

O genoma complexo do HIV-1 é constituído por duas moléculas lineares de RNA, idênticas, de cadeia simples e polaridade positiva, com cerca de 9,5 kb. Na forma proviral, este encontra-se flanqueado por duas regiões terminais não codificantes, de cerca de 600 nucleótidos, designadas por repetições terminais longas LTR (do inglês, *long terminal repeats*), com a estrutura U3RU5 (**Figura 1.7**). Nas LTR

estão localizados elementos promotores importantes que asseguram a transcrição do DNA proviral. O genoma viral é ainda constituído por uma região central codificante, que inclui os seguintes genes (**Figura 1.7**):

gag, *pol* e *env*, que codificam as proteínas estruturais do virião e enzimas (o gene *gag* codifica as proteínas estruturais MA, CA e NC e ainda a proteína p6; o gene *pol* codifica as três enzimas, cruciais no ciclo replicativo, RT, PR e IN; o gene *env* codifica as duas glicoproteínas, SU e TM);

tat e *rev*, que codificam as duas proteínas reguladoras Tat e Rev;

nef, *vif*, *vpr* e *vpu*, que codificam as quatro proteínas acessórias Nef, Vif, Vpr e Vpu.

As funções de cada uma das proteínas codificadas pelos nove genes que constituem o genoma do HIV-1 encontram-se resumidas no esquema que se segue (71, 146, 199, 209):

Genes Estruturais

<i>gag</i>	<i>pol</i>	<i>env</i>
<p>MA (p17) - Proteína estrutural da matriz; participa no transporte do complexo de pré-integração para o núcleo da célula; direcciona as poliproteínas Gag e Gag-Pol para a membrana citoplasmática; intervém na saída dos viriões;</p> <p>CA (p24) - Proteína estrutural da cápside;</p> <p>NC (p7) - Proteína estrutural da nucleocápside;</p> <p>p6^{Gag} - Proteína essencial na montagem e libertação dos viriões para o exterior da célula.</p>	<p>PR (p11) - Responsável pelo processamento das poliproteínas Pr55^{Gag} e Pr160^{Gag-Pol}, durante a fase de maturação, cuja clivagem dita a infecciosidade das partículas virais;</p> <p>RT (p66/p51) - Implicada na conversão do RNA viral de cadeia simples em DNA proviral de cadeia dupla (transcrição reversa);</p> <p>IN (p32) - Envolvida no transporte do complexo de pré-integração para o núcleo da célula e no mecanismo de integração do DNA retrotranscrito no genoma das células hospedeiras (provírus).</p>	<p>SU (gp120) - Glicoproteína de superfície, responsável pela ligação ao receptor celular CD4 e a co-receptores presentes nas células alvo (CCR5/CXCR4).</p> <p>TM (gp41) - Glicoproteína transmembranar, responsável pela fusão do invólucro viral com a membrana celular.</p>

Genes Acessórios

<i>nef</i>	<i>vif</i>	<i>vpr</i>	<i>vpu</i>
<p>Nef (p27) - Modulação negativa, entre outras, das moléculas CD4, MHC I e II à superfície da célula (escape imunitário) e aumento da infecciosidade do viriões; proteína multifuncional, importante factor de patogenicidade.</p>	<p>Vif (p23) - Promove a infecciosidade viral, impedindo a incorporação das proteínas APOBEC3F e G; contribui para a paragem do ciclo celular na fase G2.</p>	<p>Vpr (p15) - Participa no transporte do complexo de pré-integração para o núcleo da célula infectada e tem a capacidade de reter as células em divisão na fase G2 do ciclo celular; induz apoptose.</p>	<p>Vpu (p16) - Induz a degradação de CD4 no proteassoma, aumentando a capacidade de libertação das partículas virais das células infectadas.</p>

Genes Reguladores

<i>tat</i>	<i>rev</i>
<p>Tat (p14) - Aumento da actividade da RNA polimerase II celular, fundamental para a transcrição viral a partir do promotor LTR.</p>	<p>Rev (p19) - Exportação nuclear dos RNAs virais não processados, ou parcialmente processados, para o citoplasma.</p>

1.8 Ciclo Replicativo Viral

A replicação do HIV-1 ocorre preferencialmente, e de forma contínua, no tecido linfo-reticular, sendo este, não só o local da maior produção de vírus, como também o seu reservatório mais importante (24). Em seguida, encontram-se as várias etapas do ciclo replicativo viral (**Figura 1.13**):

Adsorção e entrada na célula hospedeira

O processo de infecção das células pelo HIV-1 tem início com a adsorção do seu invólucro à superfície celular, através de uma primeira ligação da glicoproteína gp120 ao receptor de membrana CD4. As mudanças conformacionais daí resultantes levam à exposição de novos locais da gp120, permitindo a sua ligação a um co-receptor (receptor transmembranar de quimiocinas), mais frequentemente, CCR5 ou CXCR4. Por sua vez, a formação do complexo gp120/CD4/co-receptor induz um rearranjo na conformação da glicoproteína gp41, permitindo a exposição de uma região hidrofóbica, denominada por péptido de

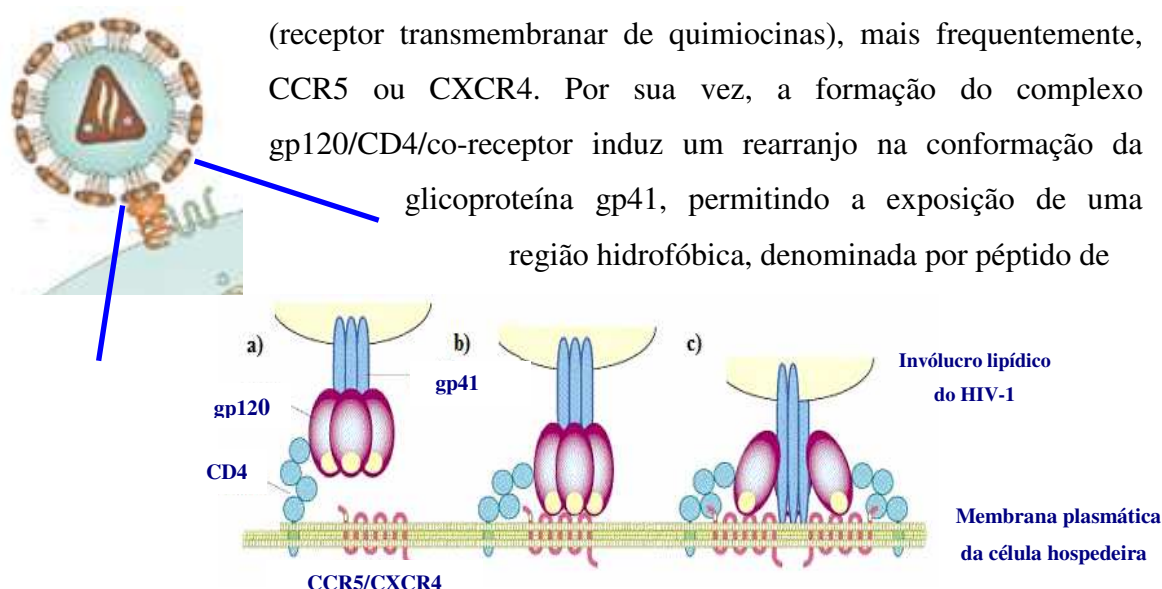


Figura 1.8. Modelo esquemático das fases iniciais da entrada do HIV-1 na célula hospedeira. a) Ligação da gp120 ao receptor membranal CD4; b) Formação do complexo gp120/CD4/receptor transmembranar de quimiocina (co-receptor); c) Exposição do péptido de fusão da gp41 e sua inserção na membrana plasmática da célula hospedeira (adaptado de 199).

fusão. Este insere-se na membrana da célula hospedeira possibilitando a sua fusão com o invólucro viral e consequente formação de um poro permissivo à entrada da cápside viral para o interior da célula (73, 146, 199) (**Figura 1.8**). Num estudo recente (136), foi proposto que a fusão viral completa não ocorre na superfície da membrana da célula hospedeira, tendo o vírus de recorrer a um mecanismo alternativo de endocitose (224). Esta descoberta implica que a próxima geração de inibidores de entrada/fusão tenha a

capacidade de passar através da membrana da célula e bloquear os eventos de fusão intracelular.

Descapsidação

Imediatamente após a sua entrada na célula hospedeira, a cápside viral é destruída por acção de enzimas celulares, processo este conhecido por descapsidação (70, 146). A ciclofilina A, uma peptidil-prolil isomerase celular, associada ao virião, parece facilitar a descapsidação (217) e posterior montagem de um complexo ribonucleoproteico de transcrição reversa (RTC, do inglês *reverse transcription complex*) funcional (73).

Transcrição reversa do genoma viral

Em simultâneo com a descapsidação, partículas subvirais são libertadas para o citoplasma da célula, onde são convertidas no complexo RTC, responsável pela síntese

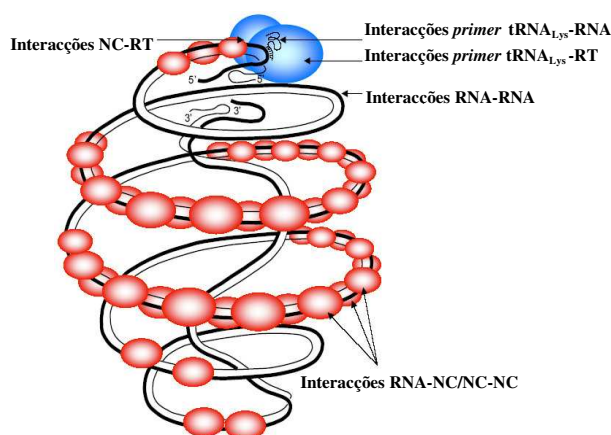


Figura 1.9. Representação esquemática do complexo ribonucleoproteico RTC e dos diferentes tipos de interações intermoleculares estabelecidos (137).

do cDNA (do inglês, *complementary DNA*) viral (17, 139). Este complexo, no qual se estabelecem vários tipos de interações moleculares (**Figura 1.9**), é constituído pelas duas moléculas de RNA genómico, revestidas pela proteína NC, em associação com outras proteínas/enzimas virais (MA, CA, Vpr, RT e IN) (139). O

início da retrotranscrição é caracterizado pela ligação do *primer* tRNA_{Lys3} celular à extremidade 5' do RNA viral, servindo de matriz para a síntese, pela RT, de uma molécula de DNA de polaridade negativa (109). Posteriormente, a molécula de DNA recém sintetizada serve como molde para a síntese de uma molécula de DNA de polaridade positiva, após degradação da molécula original de RNA. Para tal, a RT acumula três funções enzimáticas, nomeadamente, polimerase de DNA dependente de RNA, RNase H e polimerase de DNA dependente de DNA (188). No final do processo de retrotranscrição obtém-se uma molécula de DNA linear de cadeia dupla limitada pelas LTR (71).

Transporte nuclear e integração

Terminado o processo de retrotranscrição, a molécula de cDNA, juntamente com as proteínas virais RT, IN, NC, MA e Vpr (e factores celulares), faz parte do chamado complexo de pré-integração (PIC, do inglês *pre-integration complex*) (135). No citoplasma, este complexo é transportado até aos poros do invólucro nuclear, havendo possíveis interações com filamentos de actina e recrutamento de alguns factores proteicos celulares (*e.g.* LEDGF/p75), alguns dos quais envolvidos no transporte nuclear (**Figura 1.10**) (52, 166).

Uma vez no núcleo, a integrase viral, presente no PIC, catalisa a inserção do cDNA no genoma da célula hospedeira, preferencialmente em genes activos, passo fundamental no ciclo replicativo do HIV-1 (208). O cDNA integrado, denominado provírus ou DNA proviral, comporta-se agora como um “gene

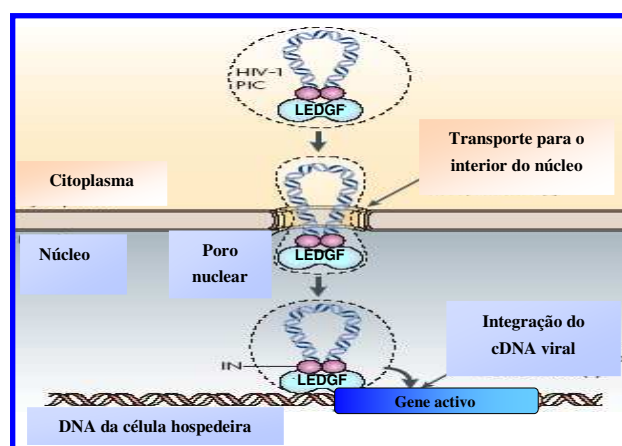


Figura 1.10. Entrada do PIC para o interior do núcleo. (adaptado de 208).

celular” (244). Dependendo do estado de activação da célula hospedeira, o provírus pode permanecer silenciado ou transcricionalmente activo, para a síntese eficiente de proteínas virais e de novas cópias de RNA genómico (208).

Num estudo recente (118), propôs-se uma nova interacção entre a proteína reguladora Rev e o co-factor celular LEDGF/p75, no contexto da integração, sugerindo-se que a formação do provírus é fortemente regulada pela interacção combinada entre ambas as proteínas e IN.

Transcrição, processamento e exportação nuclear do RNA mensageiro

A transcrição eficiente do provírus é dependente da ligação de factores de transcrição celulares, como, por exemplo, NF- κ B (do inglês, *nuclear factor κ B*), a sequências reguladoras localizadas na LTR-5' (182), direccionando posteriormente a polimerase de RNA II celular. Nos primeiros ciclos de transcrição são produzidas quantidades basais de mRNA (do inglês, *messenger RNA*) que sofrem processamento múltiplo de *splicing*, mas a expressão eficiente dos genes do provírus requer, para além

da maquinaria celular, as duas proteínas virais reguladoras Tat e Rev (73, 167). A ligação da proteína transactivadora da transcrição Tat ao elemento TAR (do inglês, *transactivation response region*), presente na extremidade 5' de todos os RNAs virais, promove, indirectamente, um aumento considerável da processividade da polimerase de RNA II celular, o que resulta na síntese de quantidades elevadas de mRNAs virais e também de moléculas completas do RNA genómico viral (24, 73). Os mRNAs sintetizados podem ser agrupados em três grandes classes (**Figura 1.11**), baseados no processamento de *splicing* sofrido (101, 167).

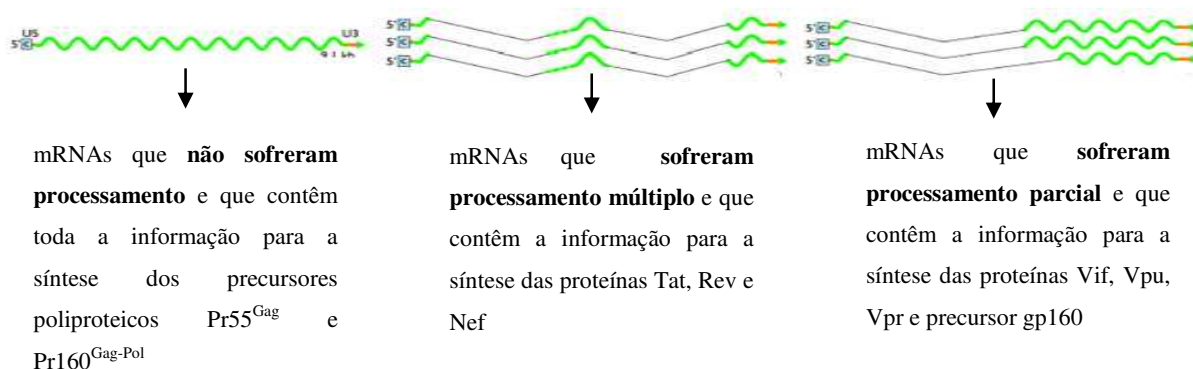


Figura 1.11. Representação esquemática das três classes de mRNAs virais sintetizados durante o ciclo replicativo, baseado no processamento de *splicing* sofrido.

O aumento da taxa da transcrição conduz à acumulação da proteína reguladora Rev, que funciona como um adaptador entre os mRNAs virais e as proteínas celulares de transporte nuclear. Na ausência desta proteína, apenas são transportados para o citoplasma os mRNAs que sofreram *splicing* múltiplo, codificando as proteínas de expressão precoce (Tat, Rev e Nef). Para levar a cabo a sua função, Rev reconhece e liga-se de forma multimérica à estrutura secundária complexa RRE (do inglês, *Rev responsive element*), presente na região *env* nos mRNAs não processados ou que sofreram um processamento parcial. Assim, a expressão precoce desta proteína permite a exportação nuclear, não só do RNA genómico viral, mas também dos mRNAs matriz para a tradução das proteínas de expressão mais tardia, codificadas pelos genes *gag*, *pol*, *env*, *vif*, *vpr* e *vpu* (73, 167).

Tradução e processamento das proteínas virais

Como se viu, as primeiras proteínas traduzidas, a partir de mRNAs que sofreram processamento completo, são as proteínas de expressão precoce Tat, Rev e Nef. Uma vez alcançados os níveis adequados, Rev permite a exportação nuclear dos restantes

transcritos, quer para os polirribossomas livres no citoplasma, onde são traduzidos, de forma massiva, dando origem a diferentes proteínas virais não glicosiladas, quer para os ribossomas associados ao retículo endoplasmático (RE), onde a poliproteína precursora gp160 é sintetizada. Durante a sua translocação, gp160 é oligomerizada, glicosilada e clivada por uma protease celular (furina), dando origem a um complexo trimérico das duas glicoproteínas do envólucro (TM/SU), associadas por ligações não covalentes (24). Em seguida, este é transportado, em associação com a membrana de vesículas secretórias, até à superfície da célula infectada, local onde já se encontram, para além dos precursores poliproteicos Pr55^{Gag} e Pr160^{Gag-Pol}, o RNA genómico e as proteínas acessórias Vif, Vpr e Nef (24, 73).

Montagem, gemulação e maturação das partículas virais

A montagem das partículas virais ocorre ao nível da membrana celular, sendo coordenada por Gag e por vários factores celulares. A libertação do virião processa-se por gemulação, processo para o qual é requerida p6^{Gag}, sendo facilitada pela proteína Vpu. Na sua ausência, a glicoproteína gp120 e o receptor CD4 tendem a formar um complexo precoce no RE, onde permanecem retidos. A interacção de Vpu com CD4 sinaliza este receptor para a via do proteassoma, onde é posteriormente degradado, permitindo que gp120 fique livre e prossiga a sua translocação até à superfície da célula. Em simultâneo com a saída, ou imediatamente depois, a partícula viral imatura inicia o processo de maturação (**Figura 1.12**), durante o qual irá sofrer um rearranjo estrutural, mediado pela actividade da PR viral. O processamento proteolítico dos precursores Pr55^{Gag} e Pr160^{Gag-Pol} dá origem a proteínas estruturais (MA, CA, NC e p6^{Gag}) e de natureza enzimática (PR, RT e IN), respectivamente, que se reorganizam para dar lugar a partículas virais maduras e infecciosas, os viriões (24, 73).

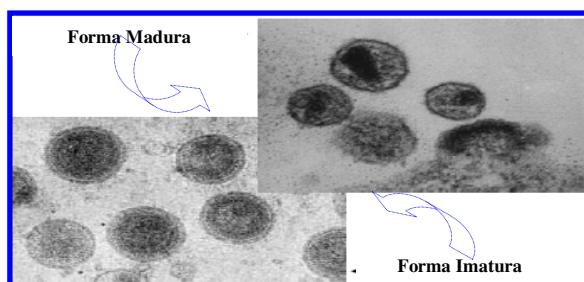


Figura 1.12. Microfotografias electrónicas das diferentes formas do HIV-1: formas madura e imatura (disponível em http://hvd.enslyon.fr/human_virology_dpt/teams/g_s_lr/pr_lr/fp_lr/2008-06_12.4108167593/image).

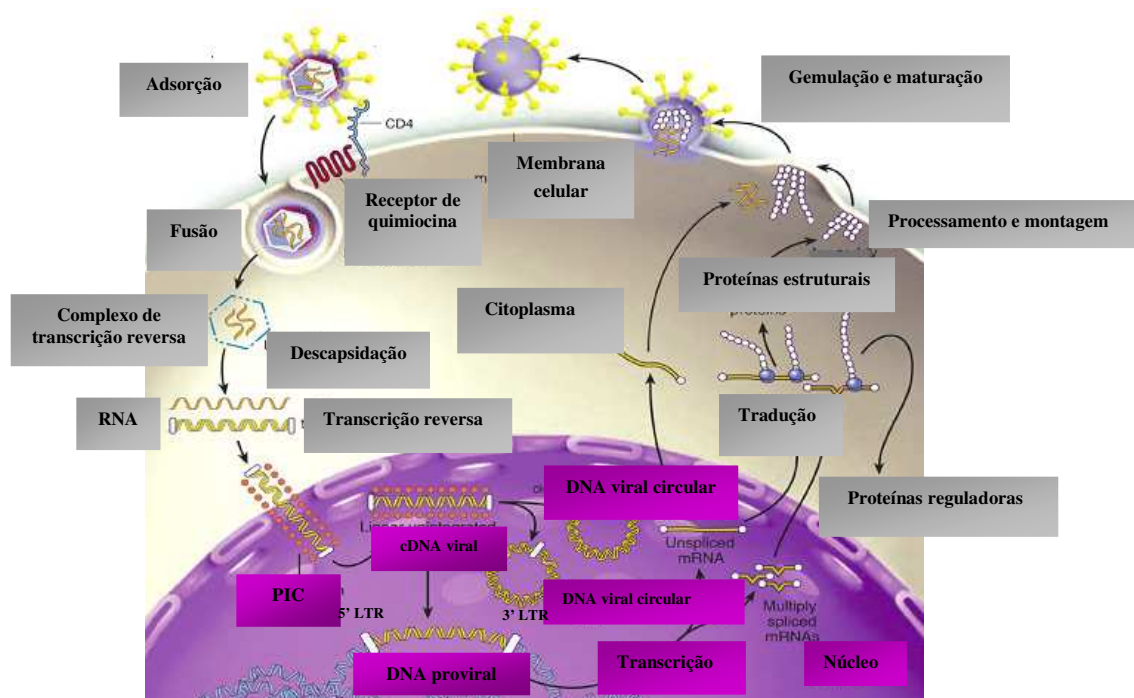


Figura 1.13. Representação esquemática do ciclo replicativo do HIV-1 (adaptado de 74).

1.9 Diversidade Genética

1.9.1 Classificação do HIV-1

O número crescente, e respectiva complexidade, de novas sequências genômicas de HIV-1 levou a uma reavaliação do sistema de classificação e nomenclatura (178). Desta forma, a análise filogenética do genoma completo de numerosas estirpes virais, com origens geográficas distintas, permitiu a classificação do HIV-1 em grupos, subtipos, sub-subtipos e respectivas formas recombinantes (178). Desta forma, os três grupos filogenéticos principais do HIV-1 são o grupo M (do inglês, *major/main*), O (do inglês, *outlier*) e N (do inglês, *non-M, non-O/new*) (178, 216). Os grupos N e O são geneticamente divergentes, entre si e relativamente ao grupo M, representando menos de 5% das infecções de HIV-1 a nível mundial (4, 180). O grupo O encontra-se praticamente confinado a indivíduos com origem na África Centro-Occidental, principalmente nos Camarões (161, 201), onde foi reportada uma prevalência que varia entre 1 a 5% (234). Para o grupo N, filogeneticamente equidistante dos outros dois, apenas se encontram descritos alguns casos pontuais de infecção, todos eles restritos à região dos Camarões (6). O grupo M inclui o grupo de vírus responsáveis pela pandemia, sendo o que engloba a maior diversidade genética (197). Actualmente (216), encontra-se dividido em linhagens filogenéticas distintas designadas por subtipos, sub-subtipos e respectivas formas recombinantes. O HIV-1 do grupo M compreende 9

subtipos virais denominados A, B, C, D, F, G, H, J e K, formando, em geral, linhagens monofiléticas consistentes, qualquer que seja a região do genoma analisada (160, 178). A variabilidade nucleotídica inter-subtipo é de cerca de 20 a 30% para o gene *env*, 15 a 22% para o gene *gag* e 10 a 12% para o gene *pol* (80, 234). Por sua vez, os subtipos A e F apresentam uma diversidade genética intra-subtipo tão elevada que foram subdivididos, respectivamente, nos sub-subtipos A1 a A4 e F1 e F2 (21, 233). Digno de nota que os subtipos B e D deveriam, de facto, corresponder a dois sub-subtipos de um mesmo subtipo genético, visto estarem mais relacionados entre si do que com os restantes (por razões históricas, esta nomenclatura é, no entanto, mantida) (43, 178). Alguns autores, através de estudos de análise filogenética, especialmente baseados nos genes *env* e *pol*, sugerem a criação dos sub-subtipos adicionais A5 e A6, de modo a acomodar novas linhagens monofiléticas divergentes dentro da radiação A (8, 235).

Recentemente, foi publicada a identificação de uma nova variante do vírus HIV-1 (165), constituindo uma linhagem filogenética distinta, evolutivamente próxima dos vírus da imunodeficiência dos gorilas (SIVgor), recentemente descobertos na espécie *Gorilla gorilla gorilla* (227). Esta nova linhagem, distinta dos outros grupos de HIV-1 (M, N e O), foi designada como protótipo de um eventual HIV-1 do grupo P.

1.9.1.1 Formas recombinantes do HIV-1 do grupo M

A co-circulação de múltiplos subtipos de HIV-1 numa área geográfica restrita favorece episódios de co-infecção (simultânea ou sequencial) de um mesmo indivíduo, o que, tendo em conta as características específicas deste vírus, conduz, muito frequentemente, ao aparecimento de vírus com moléculas de RNA genómico de origens genéticas distintas (18, 66). A existência de genomas mosaico está relacionada com o facto do vírus possuir um genoma dimérico e com a possibilidade da RT, durante a retrotranscrição, ser capaz de se dissociar do RNA matriz, reassociando-se noutra localização do mesmo ou tomando como matriz a outra molécula de RNA genómico (*template switching*) (54, 177), a qual pode ser de subtipo distinto. Do ponto de vista epidemiológico, os vírus recombinantes são designados por formas recombinantes circulantes, ou CRFs (do inglês, *circulating recombinant forms*), e formas recombinantes únicas, ou URFs (do inglês, *unique recombinant forms*). As CRFs correspondem a vírus com genomas mosaico que se propagaram com proporções epidémicas, sendo classificadas com base em sequências de genomas completos de,

pelo menos, três isolados virais, provenientes de indivíduos não relacionados do ponto de vista epidemiológico (em termos práticos, no entanto, é suficiente a sequenciação de dois genomas completos, conjuntamente com um genoma parcial que confirme a estrutura mosaico proposta). Por sua vez, as URFs correspondem a vírus com genoma mosaico identificados num único indivíduo ou num grupo de indivíduos epidemiologicamente relacionados (174, 212). Actualmente, existem descritas 45 CRFs (disponíveis em <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>), sendo responsáveis por cerca de 20% das infecções a nível mundial (113).

1.9.2 Distribuição geográfica do HIV-1 do grupo M

Hoje parece evidente que todos os subtipos do HIV-1 tiveram origem na África Centro-Occidental, numa região compreendida entre os Camarões e a República Democrática do Congo, local onde se encontram amplamente representados, corroborando a hipótese da origem e diversificação inicial do HIV na espécie humana (196). Independentemente de uma origem comum, a distribuição geográfica dos diferentes subtipos dos vírus do grupo M é bastante heterogênea (115, 174, 212), reflectindo a própria história da pandemia, vias particulares de transmissão em determinados grupos populacionais ou mesmo fenómenos de efeito fundador. O subtipo B predomina na Europa, EUA e Austrália, representando, contudo, menos de 10% do total de infecções do HIV-1 a nível mundial (115, 153, 204). Conjuntamente com as CRFs, os subtipos não-B são responsáveis por cerca de 90% das infecções, atingindo prevalências mais elevadas na África subsariana e Ásia (18, 21, 153). Independentemente das sequências do subtipo B continuarem a ser as mais representadas nas bases de dados, a nível mundial, o subtipo C é o mais frequente em todo o mundo, seguido da variante genética A (sub-subtipo A1). O primeiro encontra-se representado em quase todas as regiões do mundo, principalmente na região sul e costa leste do continente africano (18, 21). Por sua vez, o subtipo A circula, sobretudo, nos países da ex-União Soviética (em forte associação à toxicod dependência) e também na região este de África, região onde outros subtipos do HIV-1 também são encontrados, de modo que não é surpresa que tenham ocorrido fenómenos de recombinação e que muitas variantes genéticas contenham segmentos subgenómicos deste subtipo (21). Relativamente às formas virais mosaico, as que têm maior impacto na pandemia são a

CRF01_AE e a CRF02_AG, as quais circulam, predominantemente, no Sudeste Asiático e na região Centro-Oeste de África, respectivamente (18, 21, 212).

Na Europa, assiste-se actualmente ao aumento da frequência de novas infecções por variantes genéticas não-B, em relação com os fluxos migratórios oriundos de África e Ásia (64). Em Portugal predominam os subtipos G e B e formas recombinantes com origem no primeiro, CRF14_BG e CRF02_AG (**Figura 1.14**) (25, 61, 64, 65, 152, 155). Digno de realce, a necessidade de uma monitorização contínua para determinação do papel futuro dos vírus de subtipos não-B na Europa, bem como para registar o aparecimento de eventuais novos subtipos e CRFs.

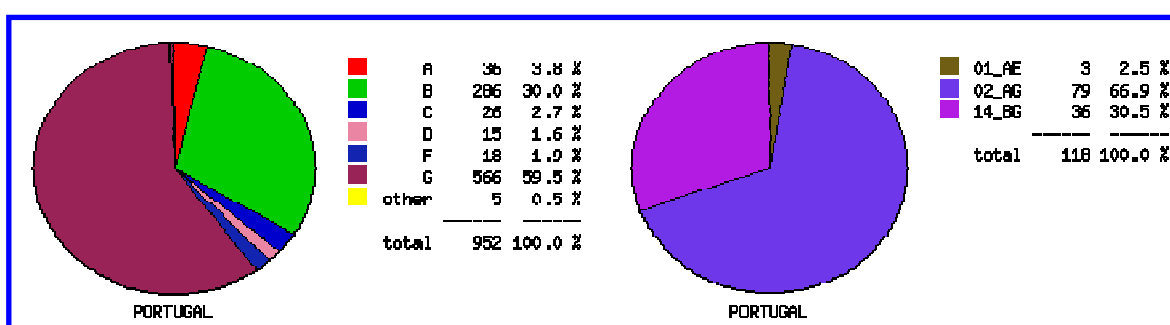


Figura 1.14. Representações gráficas da frequência das diferentes formas genéticas das sequências de HIV-1 com origem em Portugal depositadas na *Los Alamos HIV Sequence Database* (disponível em <http://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp>).

1.10 Mecanismos Responsáveis pela Diversidade Genética

A diversidade genética elevada é uma característica fundamental do HIV-1 e tem implicações importantes a vários níveis: produção de uma vacina eficaz, diagnóstico laboratorial e monitorização da infecção, transmissibilidade e patogénese, terapia anti-retroviral e epidemiologia (82, 172). Vários factores têm contribuído para a formação rápida de novas variantes genéticas, influenciando a velocidade com que estes vírus evoluem. Assim, as principais causas que contribuem para a acentuada diversidade genética apresentada pelo HIV-1 são:

Elevada taxa de incorporação de erros mediada pela RT durante a retrotranscrição do genoma viral. A incapacidade de correcção de nucleótidos erradamente introduzidos (ausência de *proofreading* ou actividade exonucleolítica 3'→5') (121) resulta numa progressiva acumulação de mutações que ocorrem a um ritmo médio estimado de, aproximadamente, 3×10^{-5} mutações por nucleótido incorporado, em cada ciclo replicativo (126).

Rápida replicação viral, estimando-se que sejam produzidas elevadas quantidades de partículas virais, aproximadamente 10^{10} partículas virais por dia/indivíduo infectado, em certas fases da história natural da infecção, com um tempo de semi-vida médio de 2,6 dias (162).

Elevada taxa de recombinação intermolecular, facultada pela forma dimérica do genoma e pela possibilidade da RT “saltar” entre as duas cadeias de RNA durante a síntese do cDNA (estima-se que a taxa de recombinação seja de três eventos de recombinação por genoma/ciclo replicativo) (102). A recombinação ocorre frequentemente em populações onde, pelo menos, duas estirpes virais geneticamente distintas co-circulam (123, 173).

A pressão selectiva elevada, exercida pelo sistema imunitário do hospedeiro e, eventualmente, pela terapêutica anti-retroviral, é responsável pela rápida selecção e fixação de formas virais com mutações de escape imune/resistência, as quais podem apresentar *fitness* elevada (115).

Os mecanismos anteriormente citados, no seu conjunto, conduzem à rápida formação de micropopulações virais distintas, geneticamente relacionadas entre si, conhecidas globalmente como *quasiespecies* (115, 177). Estas representam a “matéria-prima” fundamental para a evolução do HIV-1 (55).

A distribuição global das diferentes formas do HIV-1 é um processo dinâmico (160, 204), dependente, em última análise, do conjunto dos mecanismos responsáveis pela geração e selecção da diversidade, mas também de padrões epidemiológicos particulares, relativos às próprias populações humanas infectadas (*e.g.* fluxos migratórios, toxicodependência) (18, 21, 115). O padrão de mosaicismismo genómico, por exemplo, tende a tornar-se cada vez mais complexo, uma vez que poderá ocorrer recombinação envolvendo vírus que já são recombinantes, com a possibilidade destes vírus recombinantes de segunda geração apresentarem algumas vantagens selectivas em relação às estirpes que lhes deram origem, incluindo eventuais modificações no tropismo e na eficiência da replicação (148, 160).

1.11. Proteínas Codificadas Pelo Gene *pol*

Segundo o genoma da estirpe de referência HIV-1_{HXB2}, o gene *pol* compreende um total de 3012 nucleótidos, dos quais, 2844 correspondem à região que codifica as três proteínas de natureza enzimática PR, RT e IN (disponível em <http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2008compendium.htm>). Este gene constitui uma das regiões mais conservadas do genoma do HIV-1, o que demonstra o papel central desempenhado por estas três enzimas no ciclo replicativo viral (228). Este facto, bem como a aparente ausência de homólogos celulares, característica importante para a obtenção de baixos níveis de toxicidade na utilização de inibidores enzimáticos, faz com que estas enzimas sejam consideradas alvos atractivos da terapêutica anti-retroviral (18, 58).

1.11.1. Protease

A protease do HIV-1 pertence à família das proteases aspárticas (222). Estudos cristalográficos confirmam que esta enzima se apresenta como um homodímero

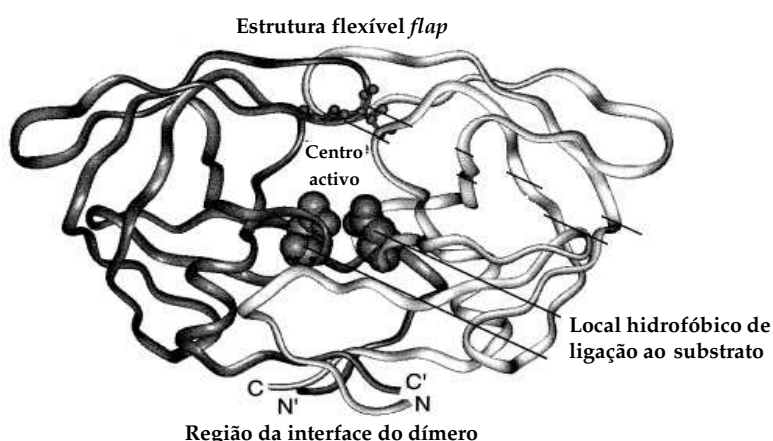


Figura 1.15. Ilustração do modelo estrutural da protease do HIV-1 (adaptado de 220).

simétrico (**Figura 1.15**), constituído por dois monómeros estruturalmente idênticos, de 99 aminoácidos cada, não ligados covalentemente, sendo que a sua actividade catalítica depende desta estrutura dimérica (239, 240). Na

sua forma activa (97), a PR possui uma estrutura flexível (*flap*, posições aminoacídicas 45 a 56) (222), que se abre e fecha sobre o centro activo (posições aminoacídicas 21 a 33), permitindo a ligação do substrato (posições 78 a 88) ou de inibidores competitivos. Como se observa na **Figura 1.15**, o centro activo da enzima localiza-se numa região central, onde se encontra uma tríade catalítica, com os resíduos D₂₅T₂₆G₂₇, altamente conservada entre as proteases dos lentivírus dos primatas (222, 240). A protease desempenha um papel fulcral no ciclo replicativo do HIV-1, visto ser necessária para o

processamento pós-traducional das poliproteínas precursoras Pr55^{Gag} e Pr160^{Gag-Pol}, as quais, após clivagem, originam as proteínas estruturais e enzimas necessárias para a correcta morfologia e infecciosidade do virião (16, 240). A inibição da actividade da protease impede o processo de maturação, dando origem a partículas virais morfologicamente defectivas e não infecciosas (97, 98). A protease possui resíduos altamente conservados (*e.g.* D₂₅, G₂₇, A₂₈, D₂₉ e G₄₉) (98), localizados principalmente no seu centro activo. Foi verificado que estes resíduos constituem bons alvos para agentes inibidores da protease, dada a sua importância para a manutenção da actividade da enzima, o que reduz, por outro lado, a probabilidade de surgirem mutantes de resistência farmacológica.

1.11.2 Transcritase reversa

A transcritase reversa é uma enzima multifuncional que tem a capacidade de converter o genoma viral de RNA de cadeia simples numa molécula de DNA de cadeia dupla (188). Apresenta-se como um heterodímero assimétrico (**Figura 1.16**) constituído por duas subunidades estruturalmente diferentes, p66 e p51, mas relacionadas, em

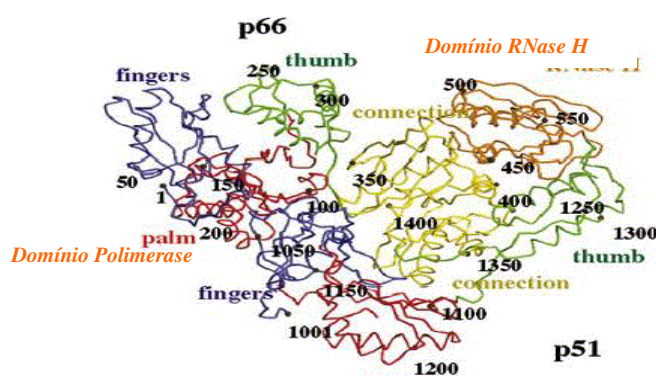


Figura 1.16. Representação esquemática da estrutura do heterodímero da RT do HIV-1 (adaptado de www.bio.davidson.edu/.../Koike/RT%20protein.htm).

termos de sequência aminoacídica (188, 202, 203). As duas subunidades partilham a extremidade N-terminal, sendo a maior (p66) constituída por 560 aminoácidos e a menor (p51) por 440 aminoácidos. Enquanto que esta última apresenta apenas um papel estrutural, a subunidade maior é detentora de dois domínios

catalíticos com funções distintas: a polimerização de cadeias de DNA, utilizando como matriz uma cadeia de RNA ou DNA (domínio polimerase), e a hidrólise do RNA genómico viral, presente no híbrido RNA/DNA, e dos *primers* tRNA_{Lys3} e PPT (do inglês, *polypurine tract*), necessários, respectivamente, para o início da síntese das cadeias de DNA de polaridade negativa e positiva (domínio RNase H) (11, 188). Estas funções enzimáticas, no seu conjunto, são essenciais para a formação de uma molécula linear de DNA de cadeia dupla (11). O domínio polimerase compreende quatro

subdomínios, cujos nomes, na nomenclatura anglo-saxónica, sugerem uma analogia com a morfologia de uma mão humana. Três deles, *fingers* (resíduos 1-85 e 118-155), *palm* (resíduos 86-117 e 156-236) e *thumb* (resíduos 237-318), são responsáveis pela polimerização, enquanto que o subdomínio *connection* (resíduos 319-426), como o nome indica, permite a ligação com o domínio RNase H (11, 188). O local de ligação do RNA viral está posicionado de modo a que este contacte directamente com o centro activo de ambos os domínios catalíticos de p66 (188). O centro activo do domínio da polimerase é constituído por uma tríade catalítica, altamente conservada, contendo resíduos de aspartato (D₁₁₀, D₁₈₅ e D₁₈₆) que formam complexos de coordenação com dois catiões divalentes, essenciais à actividade catalítica da enzima (Mn²⁺ ou Mg²⁺, sendo este último tido como o co-factor utilizado *in vivo*, dada a sua abundância relativa nas células). Os resíduos D₁₈₅ e D₁₈₆ fazem parte do motivo YXDD (X é metionina no HIV), altamente conservado em todas as RTs, sendo fundamental para manter a sua actividade de polimerase (188). Existem locais mais afastados do centro activo que possuem outros motivos/resíduos aminoacídicos conservados, igualmente importantes para manter a funcionalidade da RT. Por exemplo, o motivo estrutural *DNA primer grip* (posições aminoacídicas E₂₂₄-H₂₃₅) permite o posicionamento da extremidade 3'OH do *primer* tRNA_{Lys3} no centro activo da polimerase, iniciando-se assim o ataque nucleofílico ao α -fosfato do dNTP (do inglês, *deoxynucleoside triphosphate*) a incorporar na cadeia de DNA nascente (81, 85, 188). Foi sugerido que substituições de aminoácidos, tais como W229A, G231A e L234A, possam levar à alteração da conformação do centro activo da polimerase, originando inactivação da RT (85, 188). O local de ligação dos dNTPs inclui, por sua vez, os resíduos conservados R₇₂, K₆₅, Y₁₁₅ e Q₁₅₁, importantes na actividade catalítica da enzima, pelo que a introdução de mutações em qualquer uma das posições referidas provoca a diminuição da sua actividade (142, 188).

O domínio RNase H ocupa a região C-terminal, entre os aminoácidos 427 e 560, da subunidade p66 (179). Este domínio é detentor de duas regiões conservadas: o centro activo, constituído pelos resíduos D₄₄₂, E₄₇₈, D₄₉₈ e D₅₄₉, e a RNase H *primer grid*, no qual se destacam os aminoácidos T₄₇₃, I₅₀₅, K₄₇₆, Q₄₇₅, Y₅₀₁, Q₄₄₈, N₄₇₄ e R₅₀₀. Estudos *in vivo* e *in vitro* evidenciaram que substituições de alguns destes resíduos podem afectar a

especificidade e a actividade da RNase H, mas também da polimerase, originando vírus não infecciosos (179, 188).

Quanto à subunidade menor (p51) da RT, esta deriva da subunidade p66 por remoção proteolítica do seu domínio RNase H, por acção da protease viral. A sua função meramente estrutural é, no entanto, extremamente importante para a estabilidade e actividade do heterodímero (11, 188). Alguns estudos (81, 141, 202) sugerem que a RT se encontra activa apenas na sua forma heterodimérica, pelo que alvos terapêuticos promissores se encontram precisamente na interface entre as duas subunidades. Por outro lado, resíduos de aminoácidos altamente conservados entre as RTs dos lentivírus dos primatas, em particular, W401 e W414, são fundamentais para a formação do dímero p51/p66, demonstrando a importância da sua constituição para a actividade da enzima (143, 210). No entanto, há substituições do aminoácido 401, na subunidade p66, que embora inibam fortemente o processo de dimerização, levando à perda de actividade de polimerase, permitem manter níveis apreciáveis de actividade de RNase H (211, 237).

1.11.3 Integrase

A integrase (**Figura 1.17**) apresenta-se como um multímero, sendo formada, na sua forma activa, por dímeros ou tetrâmeros de uma cadeia polipeptídica de 288 aminoácidos, na qual se podem identificar três domínios funcionais distintos (192). Os primeiros 49 aminoácidos da proteína constituem o domínio N-terminal (DNT), no qual se destaca o motivo $H_{12}H_{16}C_{40}C_{43}$. Este promove a multimerização da integrase, para a sua forma funcional, que é caracterizada por um equilíbrio entre as formas

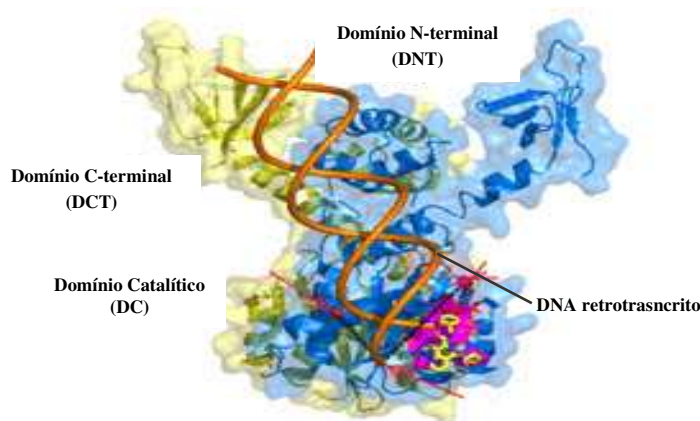


Figura 1.17. Ilustração do modelo estrutural da integrase do HIV-1, com indicação dos seus domínios funcionais (disponível em www.pnas.org/content/106/20/8192.figures-only).

dimérica e tetramérica (26). O motivo $H_{12}H_{16}C_{40}C_{43}$ forma um complexo de coordenação com um átomo de zinco, que estabiliza a proteína, sendo necessário para

sua actividade catalítica óptima (26, 52). A introdução de mutações em qualquer um dos resíduos deste motivo anula por completo a actividade da integrase (244). Entre os aminoácidos 50 e 212, está localizado o domínio catalítico (DC). Este domínio possui alguns aminoácidos bem conservados, como Q₆₂, H₆₇, N₁₂₀, N₁₁₄, Q₁₄₈ e N₁₅₅, que parecem ser essenciais para o processo de integração do DNA retrotranscrito no DNA da célula hospedeira (116). O resíduo Q148, por exemplo, é crucial na ligação entre a integrase e o cDNA viral (26), sugerindo-se que uma mutação neste resíduo levaria ao bloqueio desta interacção. O DC contém uma tríade catalítica de resíduos acídicos, que constituem o motivo D₆₄D₁₁₆E₁₅₂, altamente conservado entre as integrases dos retrovírus (26), sendo que substituições em qualquer um deles conduzem à inactivação da integrase. Outros resíduos importantes neste domínio constituem o motivo K₁₈₆R₁₈₇K₁₈₈, que participa na multimerização da enzima e que, possuindo um sinal de localização nuclear, possibilita a entrada do PIC no núcleo da célula infectada (26). Por fim, o domínio menos conservado dos três, domínio C-terminal (DCT), está compreendido entre os aminoácidos 213 e 288. O DCT está envolvido na ligação do cDNA viral ao DNA da célula hospedeira, por intermédio de uma região inespecífica (MDBD, do inglês *minimal nonspecific DNA binding region*), situada entre os aminoácidos 220 e 270. Para além de estar envolvido na etapa de integração do cDNA viral, o DCT também participa na interacção com a RT e no processo de multimerização da integrase (26).

Como se viu, a integrase é indispensável para a formação do provírus. Este processo encontra-se dividido em quatro fases distintas (26, 52, 116, 144, 229, 244), a última das quais poderá envolver a actividade de enzimas celulares:

Ligação da integrase à região da LTR do DNA retrotranscrito. Este passo é caracterizado pela formação de um complexo estável entre a forma tetramérica da integrase e o DNA, ao nível do PIC;

Processamento 3', o qual é necessário para a entrada do PIC no núcleo. Consiste, num corte endonucleolítico dos dinucleótidos 3'-TG presentes em cada extremidade das LTR do cDNA viral, adjacentes a dinucleótidos altamente conservados (AC), produzindo extremidades 3'OH livres;

Reacção de transferência de cadeia, que ocorre já ao nível do núcleo. Após a entrada, a integrase liga-se ao DNA cromossómico, mediando um ataque nucleofílico das extremidades 3'OH livres do cDNA viral ao seu esqueleto fosfodiéster. Cada uma das cadeias simples do DNA alvo sofre um corte de uma ligação fosfodiéster, separadas de cinco nucleótidos, sendo as extremidades 3'OH do cDNA viral ligadas covalentemente, de um modo irreversível, às extremidades 5'P protuberantes do primeiro. Desta reacção resultam duas bases salientes nas extremidades 5' do DNA proviral (5'AC) e dois segmentos de cadeia simples, com cinco nucleótidos cada (característico no caso do HIV), presentes nas extremidades do local de integração;

Reparação do DNA, etapa final caracterizada pela remoção dos nucleótidos não emparelhados 5'AC, preenchimento das duas regiões de cadeia simples com nucleótidos complementares e reposição das ligações fosfodiéster anteriormente destruídas, conduzindo à duplicação da região alvo da integração.

Alguns trabalhos descreveram a importância de diversos co-factores, de origem viral e celular, para o processo da integração (2, 228, 229). Embora a integrase, por si só, seja capaz de efectuar as reacções de processamento 3' e de transferência de cadeia, a sua actividade aumenta consideravelmente na presença dessas moléculas.

1.12 Terapêutica Anti-retroviral

Na ausência de uma vacina contra o HIV-1, a investigação nesta área tem incidido, nos últimos anos, principalmente na descoberta de novos compostos terapêuticos anti-retrovirais que actuam em diferentes estádios do ciclo replicativo viral (230). Actualmente, 27 agentes terapêuticos anti-HIV-1 encontram-se licenciados pela FDA (do inglês, *U.S. Food and Drug Administration*), em 32 formulações diferentes, incluindo combinações. Estes encontram-se categorizados em seis classes, de acordo com o seu mecanismo de acção, e numa categoria adicional que contém um produto de combinação multi-classe (disponível em <http://www.fda.gov/>). Essas classes incluem (107, 194):

Inibidores nucleosídicos da transcriptase reversa (NRTIs¹, do inglês *nucleoside reverse transcriptase inhibitors*);

Inibidores não nucleosídicos da transcriptase reversa (NNRTIs, do inglês *non-nucleoside reverse transcriptase inhibitors*);

Inibidores da protease (PIs, do inglês *protease inhibitors*);

Inibidor da integrase (INI, do inglês *integrase inhibitor*);

Inibidor de fusão;

Inibidor de entrada (antagonista do co-receptor CCR5).

A procura de novos agentes terapêuticos anti-retrovirais é um desafio constante, face à rápida emergência de variantes virais resistentes e multirresistentes. Alguns dos compostos mais recentes possuem mecanismos de acção únicos, que inibem o HIV em fases distintas do ciclo replicativo, permitindo a implementação de novos regimes terapêuticos combinados, caracterizados por elevadas taxa de supressão virológica (47, 140).

1.12.1 Inibidores da transcriptase reversa

1.12.1.1 Inibidores nucleosídicos

Os NRTIs são pró-fármacos que necessitam de um passo inicial de fosforilação intracelular (adição de três resíduos fosfato) para exercerem a sua função antiviral. São inibidores competitivos, sendo incorporados pela RT no DNA retrotranscrito nascente, em vez dos desoxirribonucleótidos naturais. A sua estrutura molecular impossibilita a formação da ligação fosfodiéster, necessária para a continuação da reacção de polimerização, actuando, assim, como terminadores de cadeia (**Figura 1.18**) (134, 171). Actualmente, encontram-se aprovados pela FDA dez NRTIs distintos (disponível em <http://www.fda.gov/>), havendo

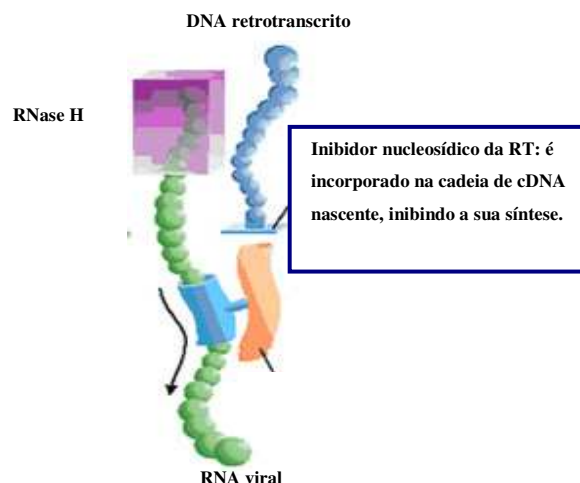


Figura 1.18. Representação esquemática do mecanismo de acção dos NRTIs (disponível em www.medscape.com/...86/458640/458640 fig.html).

¹Esta classe inclui ainda, tradicionalmente, um inibidor nucleotídico da transcriptase reversa.

comercialmente disponíveis algumas formulações combinadas. O tenofovir constitui uma excepção, uma vez que, molecularmente, é um análogo nucleotídico (a sua activação intracelular requer apenas duas etapas de fosforilação). Tradicionalmente, no entanto, é incluído nesta classe de fármacos pois o seu mecanismo de acção é semelhante ao dos NRTIs (188).

1.12.1.2 Inibidores não nucleosídicos

Os NNRTIs são pequenas moléculas de natureza hidrofóbica que inibem especificamente a RT do HIV-1 e não apresentam qualquer actividade contra o HIV-2 (171, 203). Estes inibidores actuam directamente sobre a enzima, não necessitando de

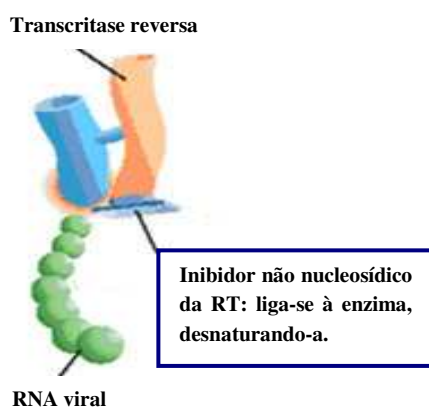


Figura 1.19. Representação esquemática do mecanismo de acção dos NNRTIs (disponível em www.medscape.com/...86/458640/458640fig.html).

fosforilação intracelular prévia para se tornarem activos. Ligam-se de forma não competitiva a um local alostérico (região hidrofóbica, fora do centro activo) da RT do HIV-1 (**Figura 1.19**) (48, 134, 203), induzindo uma alteração conformacional no seu centro activo (48). Até à data, existem quatro NNRTIs aprovados pela FDA para uso terapêutico (disponível em <http://www.fda.gov/>), sendo que um deles (delavirdine) não se encontra disponível em Portugal (231).

A existência de vários resíduos de aminoácidos conservados em diferentes regiões de ambos os domínios da RT (polimerase e RNase H) faz com que estes constituam alvos plausíveis (**Figura 1.20**) na intervenção terapêutica anti-retroviral com inibidores da RT (141, 142).

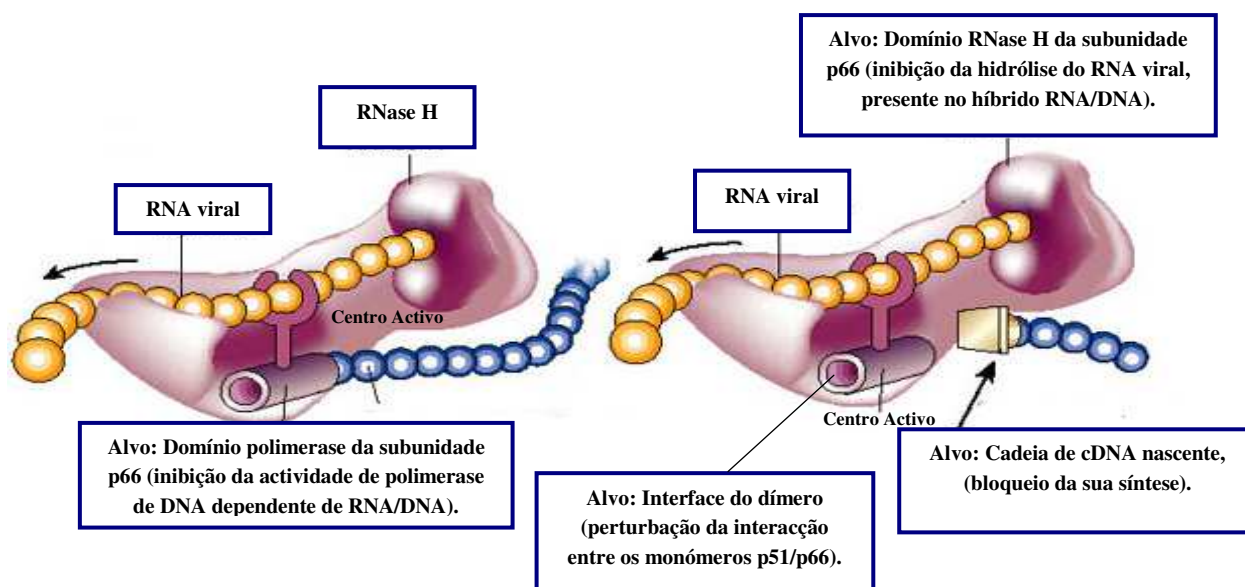


Figura 1.20. Representação esquemática de quatro alvos terapêuticos plausíveis para os inibidores da transcriptase reversa (adaptado de 176).

1.12.2 Inibidores da protease

Os PIs são inibidores competitivos que exercem a sua actividade baseada no princípio de mimetização molecular do domínio de ligação do substrato natural

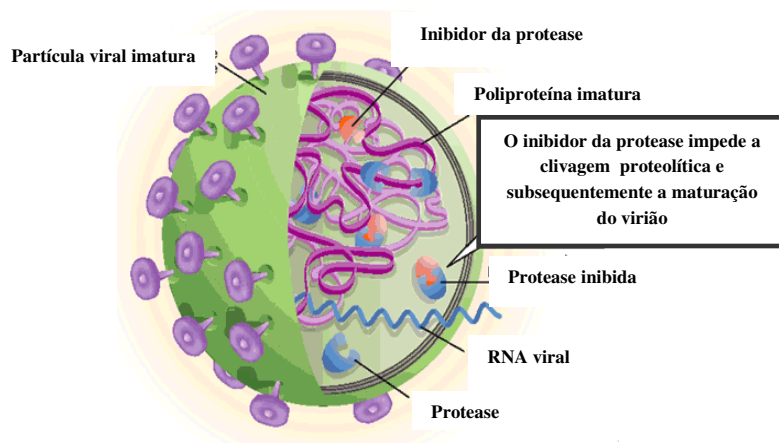


Figura 1.21. Ilustração do mecanismo de acção dos inibidores da protease (disponível em www.medscape.com/...86/458640/458640_fig.html).

(poliproteínas virais) ao centro activo da protease (48, 134). Da ligação do inibidor resulta a formação de partículas virais imaturas, estruturalmente desorganizadas e incapazes de infectar novas células (97, 110)

(Figura 1.21). São anti-

retrovirais com elevada selectividade, mas, os mais potentes, apresentam um elevado grau de ligação a proteínas séricas, causando distrofias (*e.g.* lipodistrofia, dislipidemia, intolerância à glicose, hipercolesterolemia) (215). Actualmente, são dez os PIs aprovados pela FDA (disponível em <http://www.fda.gov/>). Um destes, o ritonavir, um

inibidor do citocromo P450, é utilizado na prática clínica como potenciador farmacocinético. A administração concomitante de doses sub-terapêuticas de ritonavir com outros PIs diminui a metabolização destes, originando um aumento dos seus níveis no plasma (*PI boosting*), o que pode potenciar ou prolongar os seus efeitos terapêuticos (12, 215).

Para além dos inibidores da protease já em utilização na prática terapêutica, têm sido propostos novos agentes que bloqueiam a dimerização, ou favorecem a

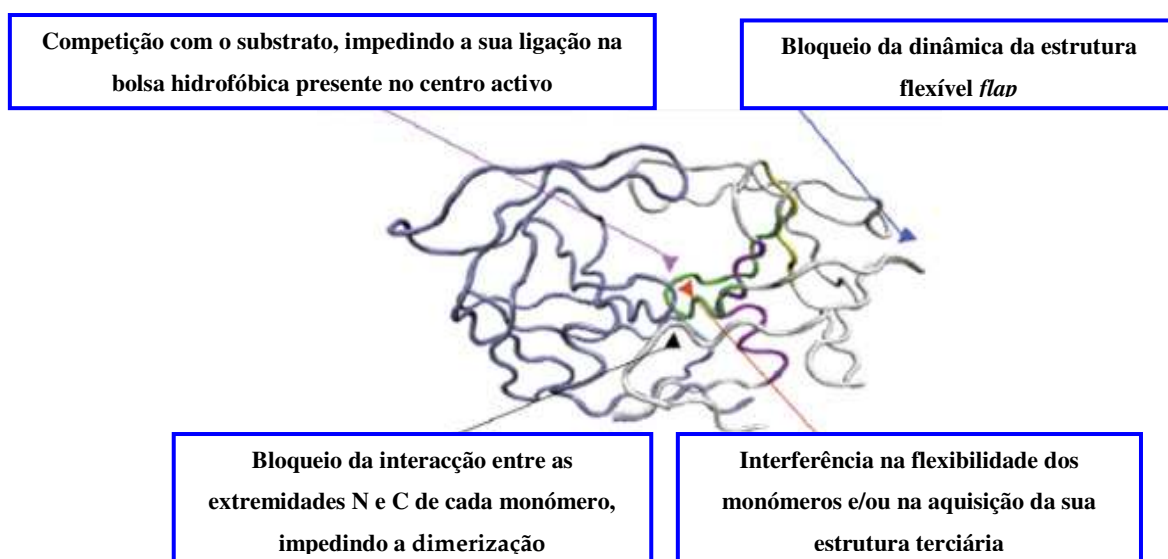


Figura 1.22. Esquema com os quatro potenciais alvos terapêuticos dos inibidores da protease (adaptado de 16).

dissociação, dos seus monómeros, contribuindo para a destabilização da forma activa da enzima (16, 98). Em alternativa, têm sido propostos inibidores que actuam na estrutura flexível *flap* e outros que interferem com a flexibilidade dos próprios monómeros da protease (222). A capacidade teórica de inibir a protease, utilizando várias abordagens (**Figura 1.22**), juntamente com a sua importante função no ciclo replicativo viral, leva a manter esta enzima como um importante alvo para o desenvolvimento de novos fármacos destinados à terapêutica anti-retroviral (97).

1.12.3. Inibidor da integrase: raltegravir

Em Outubro de 2007 foi aprovado pela FDA o primeiro, e único, até à data, inibidor da integrase, o raltegravir (69, 121, 192). Como se viu, a integração do cDNA viral no genoma do hospedeiro é essencial para a manutenção, expressão e replicação viral. Este inibidor vai bloquear (105, 134) a etapa de transferência de cadeia,

impedindo a integração, com formação de formas circularizadas do genoma viral (formas 2-LTR) (**Figura 1.23**) (90). O raltegravir é considerado um inibidor muito promissor, pelo seu forte poder inibitório e sinergismo com outros compostos anti-retrovirais (91, 192). Para além disso, apresenta elevada actividade *in vitro* contra variantes HIV multirresistentes, mantendo níveis de toxicidade aceitável (105), sendo por isso utilizado em esquemas terapêuticos de salvação (90, 124). Em Janeiro de 2008, a

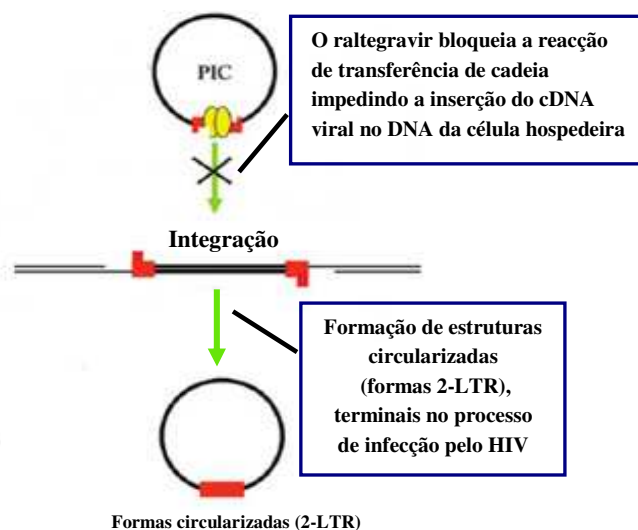


Figura 1.23. Representação esquemática do mecanismo de acção do inibidor da integrase raltegravir (adaptado de 90).

FDA determinou o uso de raltegravir em indivíduos submetidos a tratamento prévio, infectados com estirpes multirresistentes e capacidade replicativa elevada, em combinação com outros anti-retrovirais (90, 121). Em Setembro de 2009, esta indicação foi estendida para o tratamento de todos os adultos infectados com HIV-1 (previamente submetidos, ou não, à terapêutica anti-retroviral) (69, 121).

Outro inibidor da integrase, em desenvolvimento, digno de menção, é o elvitegravir (103). Este apresenta poder inibitório contra as reacções de processamento 3' e de transferência de cadeia, embora seja mais potente para esta última (105, 127). Revelou ser activo contra estirpes multirresistentes aos outros anti-retrovirais e apresenta um efeito sinérgico quando administrado em combinação (46), constituindo, portanto, um inibidor da integrase de segunda geração muito promissor (105, 192).

A ligação da integrase às LTR, em local altamente conservado entre os genomas do HIV, bem como, o passo de processamento 3' e a reacção de transferência de cadeia, são considerados importantes alvos terapêuticos (**Figura 1.24**). A inibição do processo de integração resulta, muitas vezes, na formação de estruturas circularizadas (ligação entre as LTR virais nas extremidades 5' e 3'), terminais no processo de infecção pelo HIV, sendo o vírus irreversivelmente bloqueado.

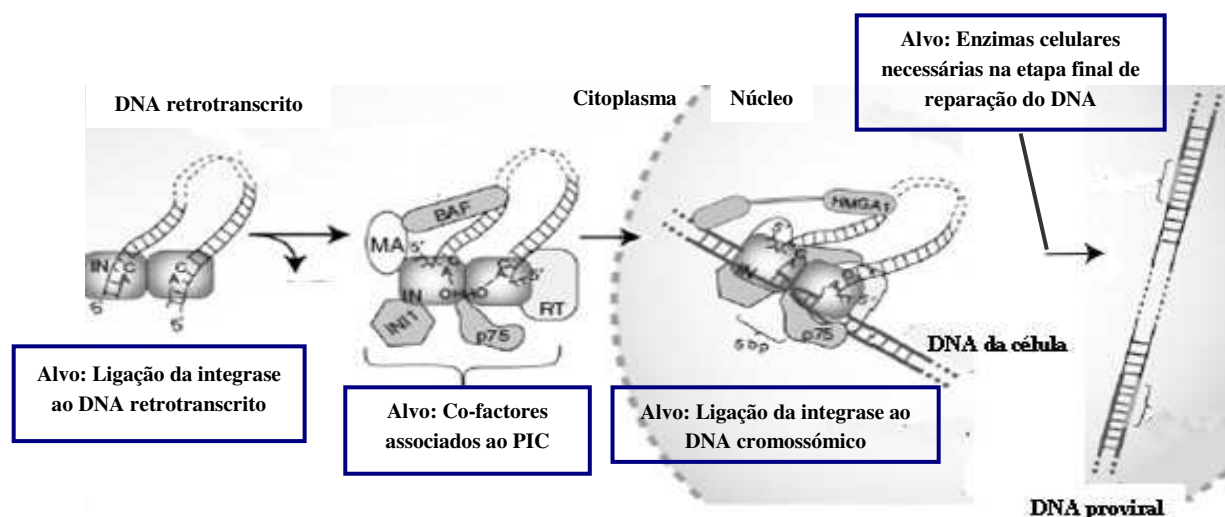


Figura 1.24. Representação esquemática de quatro alvos terapêuticos possíveis para os inibidores da integrase (adaptado de 191).

Por outro lado, entre os co-factores celulares necessários ao processo de integração, LEDGF/p75 (do inglês, *lens epithelium derived growth factor*) (20, 166) poderá constituir também um alvo terapêutico promissor, pois, permitindo a ligação da integrase à cromatina da célula hospedeira, é essencial para a integração e, consequentemente, para a replicação e infecciosidade do vírus. Na interação LEDGF/integrase participam resíduos aminoácidos altamente conservados e estudos *in vitro* demonstraram que a inibição deste co-factor bloqueia o processo de integração (26, 116).

Em suma, a caracterização de resíduos aminoácidos altamente conservados, envolvidos na estabilidade da integrase, sua multimerização, ligação ao cDNA viral, actividade catalítica, ligação a co-factores (*e.g.* LEDGF/p75), pode ajudar no desenvolvimento de novos inibidores, com mecanismos de acção alternativos aos actualmente em utilização e adequados ao perfil de resistências existente (26).

1.12.4 Terapêutica anti-retroviral de alta eficácia (HAART)

Em 1997, com o advento da terapêutica tripla, iniciou-se uma nova era no combate à SIDA, a era da terapêutica anti-retroviral de alta eficácia (HAART, do inglês *highly active antiretroviral therapy*), a qual permitiu um aumento considerável da esperança de vida dos indivíduos infectados (140, 244). A combinação de fármacos anti-retrovirais é considerada a melhor estratégia terapêutica para o controlo da infecção, na medida em que, aliada à prevenção do desenvolvimento de resistências (49), permite a supressão da replicação viral, a níveis abaixo dos limites de detecção dos testes laboratoriais de detecção de carga viral, restaurando e preservando a função imunológica, com aumento de linfócitos T CD4+ circulantes (111). A HAART, que constitui actualmente a forma de tratamento mais eficaz da infecção pelo HIV-1, consiste na administração simultânea de, pelo menos, três fármacos diferentes pertencentes, no mínimo, a duas classes distintas de anti-retrovirais (107). O esquema mais comum inclui um PI ou um NNRTI, em combinação com dois NRTIs que apresentem um efeito sinérgico. Recentemente, foi aprovado um novo regime terapêutico (raltegravir combinado com dois NNRTIs) para o tratamento de indivíduos nunca antes submetidos a terapêutica anti-retroviral (56).

No entanto, a HAART apresenta alguns inconvenientes, sendo dispendiosa, e, muitas vezes, associada a efeitos secundários que conduzem a falhas de adesão terapêutica por parte do paciente (133). Estas estão, por sua vez, na base do desenvolvimento de resistência terapêutica, o que acaba por ser, simultaneamente, causa e consequência de uma incompleta supressão do vírus, constituindo assim uma forte limitação para o sucesso da terapêutica (23, 133).

1.13 Resistência Associada à Terapêutica Anti-retroviral e Diversidade Genética

Actualmente, apesar da HAART ser genericamente eficaz no controlo da carga viral, vírus presentes em “reservatórios” do hospedeiro persistem por longos períodos de tempo, dando origem a populações virais nas quais é frequente a acumulação de variantes multirresistentes, alguns dos quais podem demonstrar grande infecciosidade e elevada capacidade replicativa (51, 244). O desenvolvimento de resistência à terapêutica anti-retroviral depende de vários factores, nomeadamente, da heterogeneidade da *quasiespecies* de HIV-1 presente no indivíduo infectado, da taxa de replicação viral apresentada durante o tratamento, da probabilidade de ocorrência de uma determinada

mutação (ou conjunto de mutações) e ainda do efeito das mutações de resistência na susceptibilidade à terapêutica e no *fitness* replicativo do vírus (13, 134). Por outro lado, têm sido frequentemente observadas mutações de resistência em vírus com origem em indivíduos nunca antes submetidos a terapia (indivíduos *naive*). Estes mutantes podem persistir durante anos como forma viral dominante e estima-se que a sua prevalência neste tipo de indivíduos possa atingir até 27% do total de formas virais presentes (128, 192).

Nos EUA e Europa, foi estimado que a prevalência de estirpes de HIV-1 resistentes aos fármacos em indivíduos com infecção adquirida recentemente era aproximadamente de 10% (8-15% nos EUA e cerca de 10% na Europa). Enquanto que noutras partes do mundo (*e.g.* África), onde o acesso à terapêutica anti-retroviral é limitado, como consequência de factores económicos e logísticos, a prevalência é inferior, podendo atingir os 5% (114). Estes números indicam claramente que as estirpes de vírus resistentes aos anti-retrovirais são passíveis de serem transmitidas entre indivíduos.

Alguns estudos demonstram que os subtipos de HIV-1 se comportam de forma diferente face à terapêutica anti-retroviral, o que torna o processo de interpretação das mutações associadas a resistência bastante complexo, ou, às vezes, mesmo controverso, sobretudo quando se trata de estirpes dos subtipos não-B, menos bem estudadas (13, 130). Já se verificou que as infecções causadas por vírus não-B são, simultaneamente, menos susceptíveis à HAART e associadas estatisticamente a uma progressão mais rápida para padrões mutacionais de resistência, quando comparados com isolados do subtipo B (204). Contudo, existem outros estudos que reportam a não existência de diferenças significativas entre subtipos B e não-B, no que se refere à eficácia da terapêutica, pelo menos de primeira linha (91). Esta observação não significa, no entanto, que os mecanismos moleculares de resistência sejam iguais para os subtipos B e não-B. Um exemplo clássico é o desenvolvimento da resistência ao nelfinavir em subtipos B e G. Vírus do subtipo B tendem a desenvolver preferencialmente as mutações de resistência D30N, L63P, V77I e/ou L90M, enquanto que indivíduos infectados com o subtipo G apresentam vírus com o perfil genotípico de resistência I54V, A71V, L89I e/ou L90M (13).

Apesar de cerca de 90% das infecções por HIV-1 a nível mundial ser devida a subtipos não-B, até agora a maior parte da informação sobre resistência a anti-retrovirais fundamenta-se principalmente em estudos realizados com vírus do subtipo B (91). Infelizmente, a maioria dos estudos sobre subtipos não-B estão pouco fundamentados e os resultados obtidos são normalmente controversos (13, 91). No entanto, o conhecimento sobre o significado evolutivo e interpretação das resistências presentes em vírus de subtipos não-B é mandatório (130). Não só pelo facto da terapêutica anti-retroviral estar a ser introduzida em países onde os subtipos não-B são claramente predominantes, mas também pelo facto do número de infecções causadas por estas variantes estar a aumentar de forma evidente na Europa. Em muitos países europeus (por exemplo, Bélgica, Reino Unido, França e Portugal), o número de indivíduos infectados por subtipos não-B é superior a 40-50% do total de novas infecções (13, 91).

1.14 Monitorização de Resistências aos Anti-retrovirais

A monitorização de resistências aos anti-retrovirais pode ser feita recorrendo-se a dois tipos de ensaio: fenotípicos e genotípicos. De uma forma genérica, os primeiros baseiam-se na obtenção prévia de vírus recombinantes contendo sequências do gene *pol*, amplificadas por RT-*nested* PCR (do inglês, *reverse transcription-nested polymerase chain reaction*) a partir do RNA genómico das estirpes em estudo, inseridas em vectores virais baseados no HIV-1, mas defectivos para aquele gene. Segue-se a medição do grau de inibição da replicação viral em sistema de culturas celulares *in vitro*, sob pressão selectiva de concentrações crescentes de diferentes fármacos anti-retrovirais, para os vírus recombinante e controlo selvagem (60). A resistência é normalmente reportada através do cálculo da concentração de fármaco capaz de inibir a replicação da estirpe em estudo em 50 ou 90% (IC50/90, do inglês *inhibitory concentration*). Os testes genotípicos, por seu lado, detectam, de um modo directo, mutações associadas à resistência fenotípica a fármacos capazes de inibir a replicação viral. Estes testes baseiam-se na sequenciação nucleotídica directa dos genes alvo, geralmente o gene *pol*, após amplificação por RT-*nested* PCR. A identificação das mutações associadas a resistência (e de outros polimorfismos genéticos) é realizada de forma semi-automatizada, através da utilização de algoritmos informáticos que permitem a elaboração de relatórios onde os padrões de mutação encontrados são associados a

diferentes níveis de susceptibilidade aos anti-retrovirais. Esta classificação depende da informação presente nas bases de dados do sistema, a qual se encontra em constante actualização por equipas internacionais de peritos (60). A interpretação dos testes genotípicos é complicada devido a vários factores: o significado clínico de muitas mutações mantém-se controverso, é difícil antever o impacto clínico da interacção entre mutações diferentes, a interpretação dos resultados genotípicos não está completamente padronizada e os ensaios actuais permitem apenas a detecção de estirpes resistentes que constituam, pelo menos, 20 a 30% do total da *quasiespecies* viral (13). Apesar de tudo, estes testes têm tido um maior impacto na prática clínica do que os testes fenotípicos, dado serem menos dispendiosos, mais rápidos e simples de executar (60).

Actualmente, para além da existência de bases de dados internacionais, em constante actualização, contendo listagens de mutações associadas a resistência [e.g. *International AIDS Society-USA (IAS-USA)*, *Los Alamos HIV Sequence Database* e *Stanford University HIV Drug Resistance Database*], encontram-se igualmente disponíveis vários algoritmos informatizados para a interpretação de dados genotípicos de resistência aos anti-retrovirais. Entre os de livre acesso, podem destacar-se os sistemas ANRS (*Agence Nationale de Recherches sur le SIDA et les Hepatites Virales*, França), HIVdb (*Stanford University HIV Drug Resistance Database*, EUA) e Rega Institute System (*Katholieke Universiteit Leuven*, Bélgica) (disponível em http://hivdb.stanford.edu/pages/asi/releaseNotes/#hivalg_algorithmsavailable). Não se baseando obrigatoriamente no mesmo tipo de premissas, estes diferentes sistemas de interpretação, quando aplicados às mesmas mutações, podem produzir resultados nem sempre coincidentes (12, 119).

1.15 Caracterização de Mutações Associadas à Resistência

As mutações associadas a resistência aos diferentes anti-retrovirais são designadas em conformidade com o fenótipo resultante, sendo por isso útil categorizá-las como principais ou *major* e acessórias ou *minor*. De uma forma geral, as primeiras reduzem por si só a susceptibilidade aos anti-retrovirais, geralmente são seleccionadas numa fase precoce do processo de acumulação de mutações de resistência, apresentam uma certa especificidade em relação ao fármaco e diminuem a capacidade de replicação do vírus. Por sua vez, as mutações acessórias são as que se acumulam no genoma do vírus que já contém, pelo menos, uma mutação principal. Estas mutações estão

associadas a efeitos compensatórios, reduzindo a susceptibilidade aos anti-retrovirais, de um modo significativo, apenas em presença de outras mutações e/ou melhorando a capacidade de replicação das estirpes já portadoras de mutações principais de resistência. No entanto, com excepções, estas definições são relativas a cada um dos fármacos, pelo que uma determinada mutação poderá tanto ser considerada principal, como acessória (193).

1.15.1 Mutações associadas à resistência aos PIs

Normalmente, para que se atinjam níveis elevados de resistência aos PIs é necessária a acumulação de mutações principais e acessórias, sendo, por isso, estes inibidores denominados de elevada barreira genética. A combinação destas mutações pode melhorar a função enzimática da protease na presença do inibidor, uma vez que a resistência pode resultar de substituições nos aminoácidos situados em regiões essenciais à sua actividade, nomeadamente no centro activo, estrutura *flap* ou no local de ligação do substrato, ou afectar ainda o seu substrato natural, provocando alterações nos locais de clivagem dos precursores polipeptídicos (13, 234). No entanto, estão descritos na literatura casos de escape terapêutico através da selecção somente de uma ou duas mutações ditas principais (*e.g.* D30N, M46I/L e L90M) (13, 194).

As mutações principais são inicialmente seleccionadas na região codificante da zona de ligação ao substrato natural e podem diferir consoante o PI administrado. Estas substituições aminoacídicas interferem com a ligação do PI à protease viral, contudo são geralmente insuficientes para inibir substancialmente a actividade antiviral do fármaco, sendo então necessárias mutações acessórias para atingir níveis clinicamente significativos na redução da susceptibilidade (13). Por outro lado, e frequentemente, a substituição de aminoácidos associada à aquisição de resistência, por exemplo, D30N, G48V, I50V, V82A e I84V, causa a diminuição da actividade catalítica da enzima e provoca um efeito deletério na capacidade replicativa do vírus resistente. Este efeito parece ser compensado pela fixação de mutações acessórias, localizadas em várias posições, que permitem a reparação da função enzimática da protease e compensam parcialmente a diminuição da capacidade replicativa do vírus (103, 171). Outro aspecto a salientar é que os PIs diferem ligeiramente na sua estrutura química, não sendo surpreendente que algumas mutações principais (*e.g.* na posição 84) originem perfis

fenotípicos de resistência múltipla (ou cruzada), constituindo este facto um dos maiores problemas da terapêutica com estes inibidores (13).

Como anteriormente mencionado, diversos tipos de estudos, em vários locais do mundo, permitem a elaboração de listagens das mutações mais frequentemente encontradas em associação com a resistência aos agentes anti-retrovirais. Na **Figura 1.25** podem consultar-se as mutações na região codificante da protease associadas à resistência aos PIs.

Atazanavir +/- ritonavir ^a	L 10 F V C	G 16 E R M I T V	K 20 M I	L 24 I	V 32 I F V	L 33 Q	E 34 I L V	M 36 I	M 46 I L	G 48 V	I 50 L	F 53 L Y V M T A	I 54 A	D 60 E	I 62 V	I 64 M V	A 71 I T L	G 73 C S T A	V 82 A T F I	I 84 V	I 85 V	N 88 S	L 90 M	I 99 L M	
Darunavir/ ritonavir ^a	V 11 I				V 32 I F	L 33 I			I 47 V		I 50 V	I 54 M L					T 74 P V	L 76 V		I 84 V		L 89 V			
Fosamprenavir/ ritonavir	L 10 F I R V				V 32 I				M 46 I L	I 47 V	I 50 V	I 54 L V M					G 73 S	L 76 V	V 82 A F S T	I 84 V			L 90 M		
Indinavir/ ritonavir ^a	L 10 I R V	K 20 M R	L 24 I		V 32 I		M 36 I		M 46 I L		I 54 V						A 71 V T	G 73 A	L 76 V I	V 82 I P T	I 84 V		L 90 M		
Lopinavir/ ritonavir ^a	L 10 F I R V	K 20 M R	L 24 I		V 32 I F	L 33 I			M 46 I L	I 47 V A	I 50 V	F 53 L V L A M T S	I 54 V				L 63 P	A 71 V T	G 73 S	L 76 V	V 82 A F T S	I 84 V		L 90 M	
Nelfinavir ^{a,u}	L 10 F I			D 30 N			M 36 I		M 46 I L								A 71 V T		V 77 I A F T S	V 82 V	I 84 V	N 88 D S	L 90 M		
Saquinavir/ ritonavir ^a	L 10 I R V		L 24 I						G 48 V		I 54 V L			I 62 V			A 71 V T	G 73 S	V 77 I A F T S	V 82 V	I 84 V		L 90 M		
Tipranavir/ ritonavir ^a	L 10 V	I 13 V	K 20 M R			L 33 F	E 35 G	M 36 I	K 43 T	M 46 I L	I 47 V		I 54 A M V	Q 58 E		H 69 K	T 74 P		V 82 L	N 83 D	I 84 V		L 90 M		

Figura 1.25. Listagem compilada pela IAS-USA das mutações principais (representadas a negrito) e acessórias na região codificante da protease do gene *pol* associadas à resistência aos inibidores da protease (103). A numeração está de acordo com a da sequência consenso do subtipo B para a região em estudo. Acima da barra encontra-se o aminoácido da forma selvagem (código de uma letra), o número corresponde à sua posição na sequência e as letras abaixo da barra indicam o aminoácido encontrado no vírus mutado, em associação com a resistência.

1.15.2 Mutações associadas à resistência aos NRTIs

Foram descritos dois mecanismos de resistência associados aos NRTIs. O primeiro é mediado pela aquisição de mutações que conferem à RT um grau de discriminação aumentado entre os dNTPs e os seus análogos, prevenindo a ligação dos inibidores durante a síntese da cadeia nascente de cDNA (*e.g.* K65R, L74V, Q151M, M184V/I). O segundo mecanismo é mediado pelo aparecimento de mutações associadas à capacidade de remoção dos NRTIs, a partir da extremidade 3' da cadeia nascente de DNA, após a sua incorporação, permitindo assim a continuação da síntese do cDNA

viral (e.g. M41L, D67N, K70R, L210W, T215Y/F, K219Q/E). Estas últimas, que estão associadas à utilização de inibidores análogos de timidina, sendo, por isso, designadas TAMs (do inglês, *thymidine analogue associated mutations*), conferem diminuição da susceptibilidade a todos os NRTIs, encontrando-se associadas a padrões de resistência cruzada (103). Os NRTIs apresentam uma barreira genética elevada, visto ser necessária uma acumulação progressiva de mutações para conferir um alto grau de resistência (**Figura 1.26**) (e.g. os complexos de mutações associadas a inserções no codão 69 e à mutação Q151M, os quais também conferem multirresistência a todos os NRTIs) (13, 194).

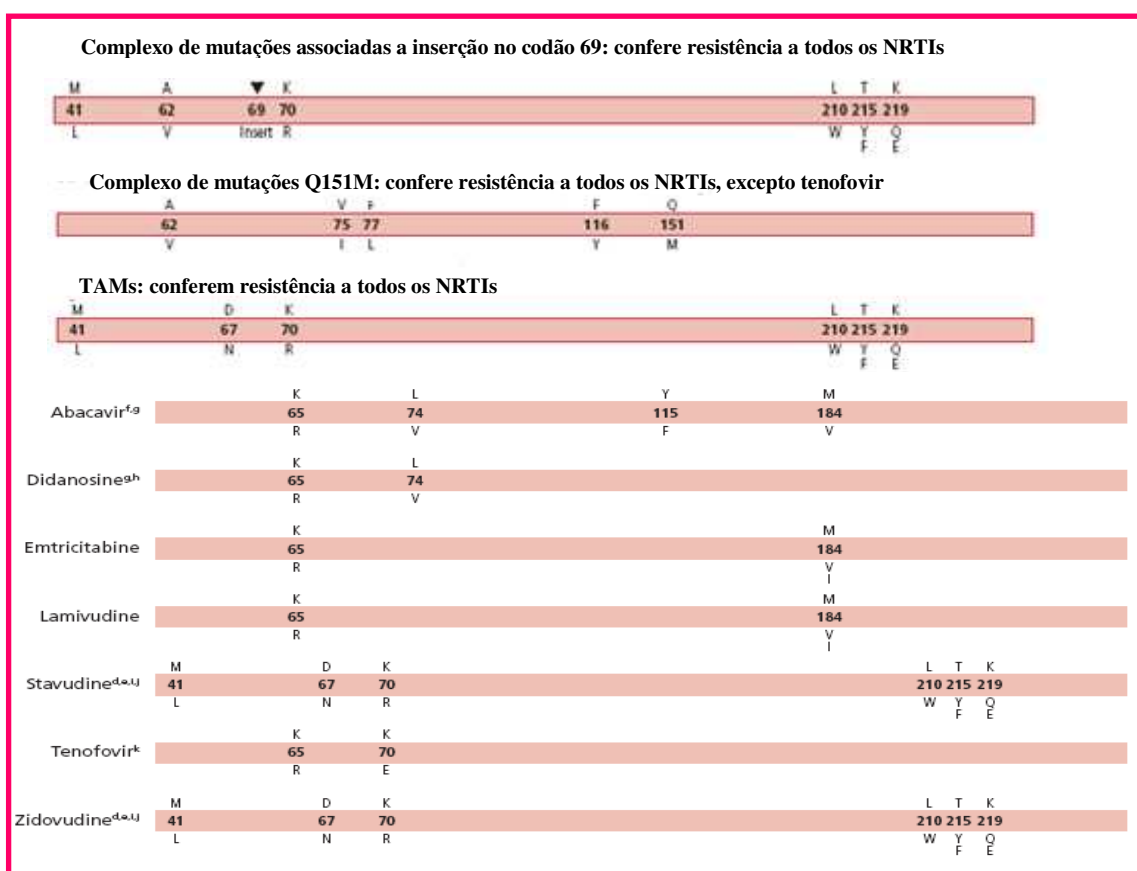


Figura 1.26 Listagem compilada pela IAS-USA das mutações na região codificante da transcriptase reversa do gene *pol* associadas à resistência aos NRTIs, incluindo as TAMs e dois complexos de mutações associadas a resistência a múltiplos NRTIs (complexos de mutações associadas a inserções no codão 69 e à mutação Q151M) (103). A numeração está de acordo com a da sequência consenso do subtipo B para a região em estudo. Acima da barra encontra-se o aminoácido da forma selvagem (código de uma letra), o número corresponde à sua posição na sequência e as letras abaixo da barra indicam o aminoácido encontrado no vírus mutado, em associação com a resistência.

1.16.3 Mutações associadas à resistência aos NNRTIs

Os NNRTIs ligam-se à RT num local de ligação específico situado próximo da tríade de resíduos catalíticos D₁₁₀D₁₈₅D₁₈₆, mas fora do centro activo da enzima. As substituições de aminoácidos associadas à resistência aos NNRTIs ocorrem ao nível desta região de ligação dos próprios inibidores (171). A acumulação de várias mutações pode ser rápida e conduzir a níveis elevados de resistência, que pode ser cruzada, para todos os fármacos desta classe. Entre as mutações de resistência aos NNRTIs listadas pela IAS-USA (**Figura 1.27**), as mais frequentes são K103N e Y181C (62). A primeira ocorre em mais de metade dos indivíduos submetidos a terapia com efavirenz e em cerca de 30% dos tratados com nevirapine. A mutação Y181C, por outro lado, tem sido encontrada em cerca de 5% dos pacientes tratados com NNRTIs (59). O uso sequencial de nevirapine e efavirenz não é recomendado, devido à emergência de resistência cruzada. Estes inibidores apresentam uma barreira genética baixa, visto a fixação de apenas uma substituição (*e.g.* K103N, V106A, G190A) ser suficiente para o vírus obter elevados níveis de resistência (13). Curiosamente, algumas mutações de resistência aos NRTIs (*e.g.* M41L, L210W, T215Y) conduzem a uma hipersusceptibilidade aos NNRTIs (103).

Esta classe de fármacos não deve ser utilizada no tratamento de infecções por HIV-1 do grupo O ou HIV-2, dada a resistência natural destes lentivírus aos NNRTIs. Níveis diferentes de susceptibilidade são também apontados para os diferentes subtipos do HIV-1 do grupo M (77).

		L	K	V	V		Y	Y	G	P
Efavirenz		100	103	106	108		181	188	190	225
		I	N	M	I		C	L	S	H
		V	A	L	K	V	E	V	Y	G
Etravirine ^m		90	98	100	101	106	138	179	181	190
		I	G	I	E	I	A	D	C	S
					H			F	I	A
					P			T	V	
		L	K	V	V		Y	Y	G	
Nevirapine		100	103	106	108		181	188	190	
		I	N	A	I		C	C	A	
				M			I	L	H	

Figura 1.27. Listagem compilada pela IAS-USA das mutações na região codificante da transcriptase reversa do gene *pol* associadas à resistência aos NNRTIs. As mutações que estão representadas a negrito conferem elevado grau de resistência (103). A numeração está de acordo com a da sequência consenso do subtipo B para a região em estudo. Acima da barra encontra-se o aminoácido da forma selvagem (código de uma letra), o número corresponde à sua posição na sequência e as letras abaixo da barra indicam o aminoácido encontrado no vírus mutado, em associação com a resistência.

1.15.4 Mutações associadas à resistência ao raltegravir

A resistência ao único inibidor da integrase licenciado (raltegravir) surge da selecção de mutações que conduzem à alteração do seu alvo terapêutico, como sejam, alterações conformacionais no DC da enzima ou perturbações na coordenação dos catiões metálicos divalentes, levando a uma reduzida afinidade do inibidor em causa para a integrase (26).

Segundo a IAS-USA, as mutações que contribuem para a resistência ao raltegravir (**Figura 1.28**) agrupam-se em combinações específicas de mutações principais (seleccionadas em primeiro lugar, na presença do inibidor, e normalmente associadas a decréscimos significativos na susceptibilidade viral) e mutações acessórias. As mutações primárias mais frequentes são Q148H/K/R e N155H (134), sendo que o padrão de resistência mais comum é a combinação Q148H+G140S, a qual confere um decréscimo de 100% da susceptibilidade ao raltegravir, mantendo o vírus uma elevada capacidade replicativa *in vitro* (103, 124, 134). Na ausência das mutações principais, o aparecimento de mutações acessórias possui um efeito mínimo na susceptibilidade *in vitro* ao raltegravir, sugerindo que estas, como o nome indica, possuem um papel acessório, com reflexo no aumento da capacidade replicativa viral ou no nível de resistência (26). Alguns estudos definem o raltegravir como um inibidor de baixa barreira genética (116, 125), ou seja, é necessário somente a presença de uma única mutação para que o vírus se torne resistente (**Figura 1.28**). Desta forma, sempre que possível, o raltegravir deveria ser administrado no contexto da HAART, de modo a minimizar o potencial para o desenvolvimento de resistência e consequente falência virológica. Contudo, outros autores refutam esta hipótese, afirmando que este inibidor apresenta uma elevada barreira genética, sendo por isso necessária a presença de várias mutações para que o vírus se torne resistente (46).

	Y	Q	N
Raltegravir ^a	143	148	155
	R H C	H K R	H

Figura 1.28. Listagem compilada pela IAS-USA das mutações principais na região codificante da integrase do gene *pol* associadas a resistência ao raltegravir (103). A numeração da sequência está de acordo com a sequência consenso do subtipo B para a região em estudo. Acima da barra encontra-se o aminoácido da forma selvagem (código de uma letra), o número corresponde à sua posição na sequência e as letras abaixo da barra indicam o aminoácido encontrado no vírus mutado, em associação com a resistência.

1.16 Objectivos

Os objectivos principais deste estudo consistiam em:

caracterizar a diversidade genética das regiões codificantes da protease (PR), transcriptase reversa (RT) e integrase (IN), do gene *pol* em estirpes de HIV-1 circulantes numa população de toxicodependentes por via endovenosa da área da Grande Lisboa;

identificar os polimorfismos genéticos presentes nessas regiões do genoma viral, associados a diferentes níveis de diminuição da susceptibilidade aos anti-retrovirais inibidores enzimáticos já licenciados ou em vias de licenciamento;

correlacionar a presença de polimorfismos genéticos específicos (e respectivo perfil fenotípico) com os diferentes subtipos virais em circulação na população estudada.

Para responder aos objectivos gerais enunciados, estabeleceram-se os seguintes objectivos específicos:

amplificar por RT-*nested* PCR as três regiões codificantes (PR, RT e IN) do gene *pol*;

clonar, se necessário, os fragmentos de DNA amplificados em vector plasmídico, para posterior sequenciação nucleotídica;

obter as sequências nucleotídicas dos diferentes amplicões, para posterior edição e alinhamento múltiplo com sequências de referência correspondentes do HIV-1 presentes em bases de dados internacionais;

genotipar as sequências PR, RT e IN, através de análise filogenética automática e “manual”;

inferir acerca da existência de formas virais recombinantes intragénicas;

pesquisar a presença de polimorfismos genéticos (definição do perfil genotípico), por comparação das sequências obtidas com sequências consenso do subtipo B;

associar aos diferentes perfis genotípicos encontrados, os respectivos perfis fenotípicos potenciais de resistência aos anti-retrovirais inibidores enzimáticos.

2. Material e Métodos

O presente trabalho encontra-se inserido num projecto de investigação envolvendo a caracterização de vírus de transmissão parentérica em circulação nos utilizadores de drogas por via endovenosa da Grande Lisboa.

O equipamento e materiais utilizados são de uso corrente num laboratório de Virologia/Biologia Molecular. Todo o material e reagentes foram esterilizados por métodos padrão de esterilização, quando necessário. As estirpes bacterianas de interesse e clones recombinantes foram conservados a -80°C em meio de crescimento LB (do inglês, *lysogeny broth*) com meio de criopreservação (1:1). A composição de todas as soluções usadas encontra-se indicada no **Anexo 2**.

2.1 Caracterização Geral da População Estudada

Este estudo envolveu um total de 51 amostras de plasma de indivíduos seropositivos para HIV-1 e utilizadores de drogas por via endovenosa. Foi preenchido voluntariamente um inquérito anónimo que permitiu obter informação sobre uma série de características epidemiológicas e clínicas (idade, sexo e país de origem dos indivíduos infectados, ano de infecção, origem dos parceiros sexuais, prática de relações sexuais não protegidas e de partilha de seringas, carga viral, estágio da infecção e terapêutica anti-retroviral), encontrando-se estas resumidas no **Anexo 3**. Genericamente, 15 (29,4%) indivíduos pertenciam ao sexo feminino e 36 (70,6%) ao sexo masculino, com idades compreendidas entre os 24 e os 49 anos (a idade média da população estudada é de $37,5 \pm 5,68$ anos).

2.1.1 Origem das amostras biológicas

As amostras biológicas estudadas foram colhidas, após obtenção de consentimento informado, entre 1998 e 2009, no:

Centro de Atendimento a Toxicodependentes das Taipas, Instituto da Droga e da Toxicodependência (CAT das Taipas, IDT)

20 amostras (TP): PT_TP02, PT_TP03, PT_TP04, PT_TP08, PT_TP13, PT_TP16, PT_TP19, PT_TP24, PT_TP38, PT_TP42, PT_TP44, PT_TP47, PT_TP55, PT_TP63, PT_TP71, PT_TP72, PT_TP85, PT_TP98, PT_TP107 e PT_TP114;

**Laboratório de Referência da SIDA, Departamento de Doenças Infecciosas,
Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA)**

30 amostras (TX): PT_TX08, PT_TX12, PT_TX13, PT_TX14, PT_TX15, PT_TX16, PT_TX18, PT_TX20, PT_TX21, PT_TX23, PT_TX27, PT_TX30, PT_TX32, PT_TX33, PT_TX34, PT_TX37, PT_TX40, PT_TX41, PT_TX43, PT_TX45, PT_TX46, PT_TX47, PT_TX48, PT_TX50, PT_TX51, PT_TX52, PT_TX54, PT_TX59, PT_TX60 e PT_TX69 (estas amostras foram gentilmente cedidas pela Doutora Elizabeth Pádua);

Hospital Prof. Doutor Fernando Fonseca, EPE (HFF)

1 amostra (AS): PT_AS646.

2.2 Processamento das Amostras Biológicas

Amostras de 10 ml de sangue periférico, colhidas por punção venosa, foram adicionadas a 0,5 ml de uma solução de citrato de sódio a 3,8% (p/v), como anticoagulante. Após a receção no laboratório, seguiu-se um passo de centrifugação durante 15 minutos, a 1800 rpm (*GPR centrifuge*, Beckman, EUA), para separação do plasma¹. Numa câmara de fluxo laminar (*Class II Type B3 Bio-Hazard Cabinet*, ENVIRCO, EUA), localizada no laboratório de nível de segurança 3 (segundo a classificação da OMS), o plasma recolhido por centrifugação (cerca de 5 ml) foi fraccionado em alíquotas de 1 ml e colocado em microtubos de criopreservação, devidamente identificados, para armazenamento a -80°C. Da fracção restante do sangue, foram isoladas e purificadas as células mononucleadas de sangue periférico, por centrifugação em gradiente de densidade, para outros estudos no laboratório, seguindo-se o protocolo descrito em Braga (15).

2.3 Pesquisa de Anticorpos Anti-HIV

O estatuto serológico para o HIV dos indivíduos do CAT das Taipas e do HFF era fornecido *a priori*, mas foi confirmado por um teste imunocromatográfico rápido (*Hexagon HIV*, Human, Alemanha), realizado de acordo com as especificações do fabricante. O plasma foi testado para a presença de anticorpos anti-HIV-1 e anti-HIV-2,

¹ As amostras provenientes do INSA foram fornecidas já com a fracção correspondente de plasma, tendo-se armazenado a -80°C, logo após a sua receção no laboratório.

tendo-se considerado positivo após visualização de reactividade contra as proteínas virais gp41/p24 ou gp36, respectivamente². As amostras com um resultado serológico negativo foram descartadas deste estudo.

2.4 Extracção do RNA Viral a Partir de Plasma

A extracção de RNA viral, a partir de plasma, foi realizada recorrendo ao sistema *INSTANT Virus RNA Kit* (Analytik Jena AG, Alemanha), segundo as especificações do fabricante. Iniciou-se o protocolo pela adição de 450 µl de solução de lise a tubos de extracção devidamente identificados. Num compartimento físico separado, em câmara de fluxo laminar (*Class II Type B3 Bio-Hazard Cabinet*, ENVIRCO, EUA) foram adicionados 150 µl das respectivas amostras de plasma. De seguida, os conteúdos dos tubos foram homogeneizados e permaneceram durante 15 minutos, à temperatura ambiente, período durante o qual foram aplicados três a quatro pulsos de vórtex. Nesta etapa, procede-se à disrupção dos viriões presentes nas amostras de plasma, com protecção do RNA viral da degradação pelas RNases. Terminado o passo anterior, os tubos foram rapidamente centrifugados (*112 centrifuge*, Sigma, Alemanha), apenas para remover o condensado das tampas, sendo depois adicionado um volume de 600 µl de solução de ligação (com o propósito de fornecer as condições necessárias para a ligação do RNA à coluna). Após mistura completa no vórtex, as colunas (*spins filters*) foram aplicadas a tubos colectores, devidamente identificados, aos quais se adicionaram as respectivas amostras (em duas vezes, consecutivas). Seguiu-se um passo de centrifugação a 12000 rpm, durante 1 minuto (*112 centrifuge*, Sigma, Alemanha), e, posteriormente, a lavagem do RNA retido nos *spin filters*, com o objectivo de eliminar todas as ligações inespecíficas presentes. O processo de lavagem foi realizado, em duas fases, com 500 e 650 µl das soluções de lavagem HS (do inglês, *high salt*) e LS (do inglês, *low salt*), respectivamente, a 12000 rpm, durante 1 minuto (*112 centrifuge*, Sigma, Alemanha). Após um último passo de centrifugação a 12000 rpm, durante 2 minutos (*112 centrifuge*, Sigma, Alemanha), para eliminação total do etanol presente nas soluções de lavagem, os *spins filters* foram transferidos para tubos de eluição, devidamente identificados, tendo-se adicionado 30 µl de água sem RNases. Os tubos permaneceram durante 2 minutos à temperatura ambiente e, de seguida, centrifugados a

²É de referir que as amostras fornecidas pelo INSA não foram testadas para a presença de anticorpos anti-HIV-1, visto já terem sido alvo de estudo para outras regiões do genoma viral, nomeadamente *gag* e *env*.

12000 rpm, 1 minuto (*112 centrifuge*, Sigma, Alemanha), sendo este passo de eluição repetido duas vezes. No cômputo geral, o volume eluído foi distribuído por cinco microtubos livres de RNases e DNases de 0,5 ml (*Multiply[®]-Pro*, Sarstedt, Alemanha), três dos quais continham o eluato mais concentrado, com um volume final de 10 µl, e os dois restantes, o menos concentrado, correspondente à segunda eluição (15 µl). As amostras de RNA foram conservadas a -80°C. Para a realização deste protocolo, foram seguidos alguns procedimentos genéricos adaptados ao manuseamento de RNA, dada a elevada susceptibilidade desta molécula orgânica à degradação por RNases endógenas ou ambientais (cumpriram-se as recomendações do sistema *INSTANT Virus RNA*, Analytik Jena AG).

2.5. Origem e Preparação dos *Primers*

Os oligonucleótidos sintéticos iniciadores (ou *primers*) utilizados para a amplificação das regiões codificantes da transcriptase reversa (RT) e da protease (PR), a partir do RNA viral, por RT-*nested* PCR (do inglês, *reverse transcription-nested polymerase chain reaction*), foram adaptados de Steegen *et al.* (207) (**Tabela 2.1**). Com o objectivo de verificar se os *primers* descritos apresentavam igual especificidade para os diferentes subtipos de HIV-1, fez-se uma comparação das respectivas sequências contra um alinhamento múltiplo de sequências nucleotídicas de estirpes de referência HIV-1 do grupo M, de vários subtipos, disponíveis na base de dados *Los Alamos HIV database* (em <http://hiv-web.lanl.gov/>). Após inspecção visual do alinhamento efectuado, foram alteradas, pontualmente, algumas das sequências nucleotídicas dos *primers* descritos por Steegen *et al.* (207) (Gag2, PR1, RT137, PR3, PR4 e RT4), com o propósito de aumentar ainda mais a sua especificidade (R. Parreira, com. pessoal). Por sua vez, os *primers* usados nas reacções de transcrição reversa e amplificação da região codificante da integrase (IN) são os descritos por Braga (15). Os *primers* directos (**Tabela 2.1**) foram desenhados com base numa sequência consenso de subtipos, sub-subtipos e CRFs representativas do HIV-1 (disponível em <http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2005compendium.htm>), para a região do genoma em causa, enquanto que os *primers* inversos foram adaptados de Burns *et al.* (19). Alguns dos *primers* utilizados neste estudo (PR1, RT137, RT3303, RT4, polOc_for, polO_for e polI_for), (**Tabela 2.1**) apresentam degenerações (Y=T/C, R=G/A, W=A/T), nas posições para as quais se verificou maior

variabilidade entre as sequências alinhadas. As sequências e coordenadas dos *primers* estão indicadas na **Tabela 2.1** e a numeração dos nucleótidos encontra-se atribuída com base na estirpe de referência HXB2 do subtipo B de HIV-1.

Todos os *primers* utilizados neste estudo foram sintetizados e purificados na empresa Eurofins MWG Operon (Alemanha). A sua preparação, a partir do liofilizado, foi realizada em câmara de fluxo laminar (*Laminar Flow Cabinet-class 100*, Flow Laboratories, EUA), utilizando-se pontas de micropipeta com filtro *10 REACH™* (Molecular Bioproducts, EUA). Cada um dos *primers*, referentes às regiões PR/RT e IN, foi hidratado com água ultrapura *Milli-Q®* (*Reagent Grade Ultrapure Systems*, Millipore, EUA), para uma concentração final de 200 pmol/μl e de 100 pmol/μl, respectivamente. Posteriormente, os *stocks* de *primers* foram diluídos para soluções de trabalho a 10 pmol/μl, mantidas a -20°C.

Tabela 2.1. Primers utilizados para a 1ª e 2ª reacções de RT-nested PCR

RT-nested PCR	Região amplificada	Nome do <i>primer</i>	Posição em HXB2 (pb)	Sequência 5´- 3´	
<i>Primers Externos</i> (RT/1ª Reacção)	Protease & Transcritase reversa	Gag2*	1404-1426	CAATGAGGAAGCTGCAGAATGGG	
		PR1*	1822-1844	GAAATGATGACAGCATGYCAGG G	
		RT137*	3301-3325	TTCTGTATRTCATTGACAGTCCAGC	
		RT3303*	3309-3329	TAA YTTYTGATRTCATTGAC	
	Integrase	polOc_for [#]	3855-3877	GGAGCAGARACYTWYTATGTAGA	
		polO_for [#]	4059-4080	GGAATCATT CARGCACAACCAG	
		HIV-vif2_rev [§]	5243-5264	CCTGTATGCAGACCCCAATATG	
	<i>Primers Internos</i> (2ª Reacção)	Protease	PR3*	2146-2166	CCAGAGCCAACAGCCCCACCA
			PR4*	2584-2605	GGGCCATCCATTCTGGCTTTA
		Transcritase reversa	RT1*	2604-2632	CCAAAAGTTAAACAATGGCCATTGACAGA
RT4*			3231-3253	TGGAGTTCATAWCCCATCCAAAG	
Integrase		polI_for [#]	4159-4180	CAGCACACAAAGGRATTGGAGG	
		HIV-pol8_rev [§]	5043-5064	CATCACCTGCCATCTGTTTTC	

*Adaptados de Steegen *et al.* (207); [#]adaptados de Braga (15); [§]adaptados de Burns *et al.* (19)

2.6 Retrotranscrição e Amplificação de Regiões do Gene *pol*

Com o intuito de se iniciar a sua caracterização genética, foram amplificadas, por RT-nested PCR, as três regiões codificantes do gene *pol*, nomeadamente, PR, RT e IN, do HIV-1, com utilização do sistema *Titan One Tube RT-PCR* (Roche Applied Science,

Alemanha). Este sistema é altamente sensível e permite associar, num único passo, a reacção de conversão do RNA viral numa molécula de cDNA (do inglês,

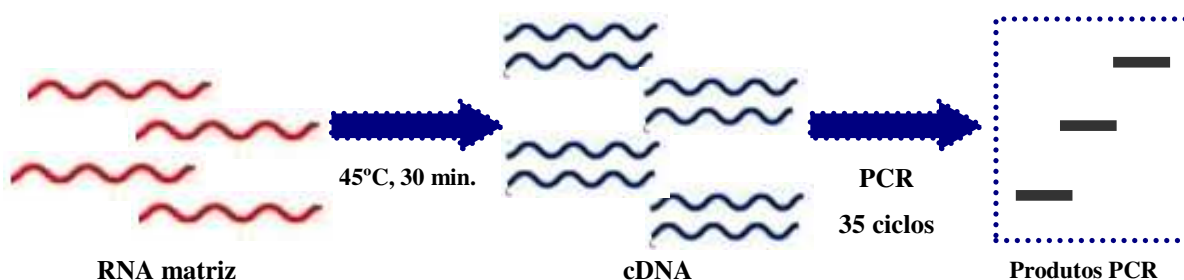


Figura 2.1. Representação esquemática geral do princípio do sistema *Titan One Tube RT-PCR* (Roche Applied Science, Alemanha).

complementary DNA), pela acção da RT do vírus da mieloblastose aviária (AMV, do inglês *avian myeloblastosis virus*), e a sua amplificação subsequente, através da combinação de duas enzimas, as polimerases de DNA Taq e Tgo, cuja combinação permite o aumento da fidelidade da PCR (Figura 2.1).

Neste estudo, foi planeada uma RT-*nested* PCR, sendo que os fragmentos de DNA obtidos na primeira reacção de amplificação, utilizando o sistema descrito anteriormente, serviram como matriz para a realização de uma segunda reacção de amplificação. Os *primers* utilizados na primeira reacção vão hibridar com uma região externa daquela que se pretende estudar, sendo por isso designados de *primers* externos. Para a segunda reacção de PCR, os *primers* internos permitiram a amplificação de fragmentos de DNA correspondentes a cada uma das regiões codificantes do gene *pol* em estudo (PR, RT e IN).

2.6.1 Planificação do protocolo de amplificação por RT-*nested* PCR

A planificação do protocolo de amplificação foi realizada com base nos estudos anteriormente referidos (15, 207), tendo-se sempre em conta as temperaturas de desnaturação dos *primers*, o grau de degeneração de alguns deles e as dimensões teóricas do fragmento de DNA a amplificar. Para se proceder, em simultâneo, à retrotranscrição do RNA viral e à primeira amplificação por PCR das regiões PR e RT, por RT-PCR, foram usados quatro *primers* externos, dois directos (Gag2, PR1) e dois inversos (RT137, RT3303) (Tabela 2.1 e Figura 2.2), de acordo com o protocolo descrito por Steegen *et al.* (207), permitindo a amplificação de quatro fragmentos de cDNA *gag-pol* distintos, de 1926, 1922, 1508 e 1504 pares de bases (pb). Estes

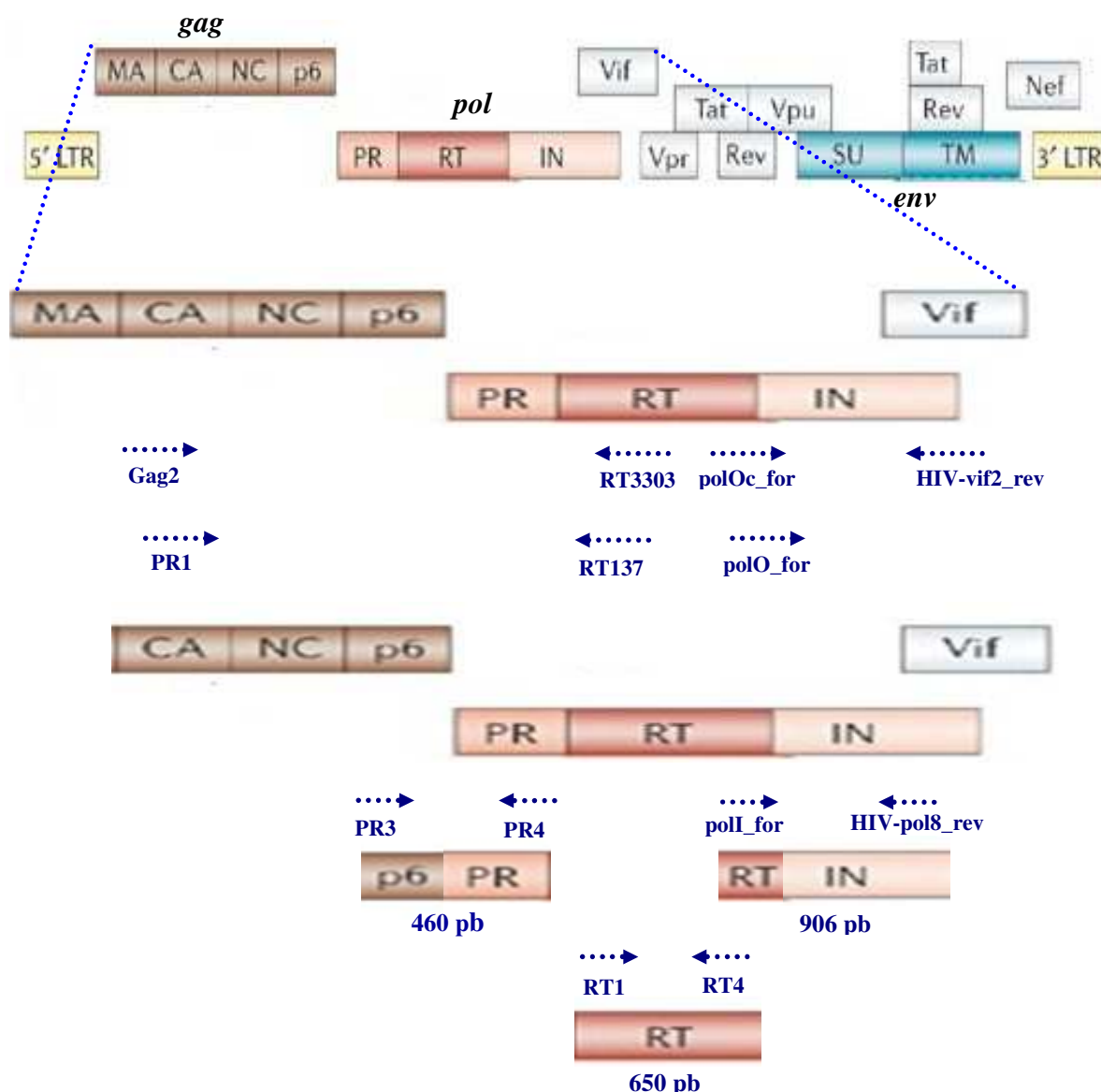


Figura 2.2. Representação esquemática da amplificação das regiões da protease (PR), transcriptase reversa (RT) e integrase (IN) do gene *pol* do HIV-1, por RT-nested PCR, para a obtenção de fragmentos de 460, 650 e 906 pb, respectivamente (adaptado de www.nature.com/./fig_tab/nrmicro1477_F1.html). A representação não se encontra à escala.

serviram como matrizes para uma segunda reacção de PCR para amplificação, em separado, de fragmentos de DNA das regiões PR e RT, com os pares de *primers* PR3/PR4 e RT1/RT4, respectivamente (Tabela 2.1 e Figura 2.2). A hibridação de PR3 ocorre numa região do gene *gag* codificante para a proteína p6 e PR4 hibrida numa região do gene *pol* codificante para a actividade da RT (p51), permitindo a amplificação de um fragmento de DNA com 460 pb (Figura 2.2). Quanto aos *primers* para a amplificação da RT, estes vão hibridar na sua região codificante, no gene *pol*,

permitindo a amplificação de um fragmento de DNA de 650 pb (**Figura 2.2**). Posteriormente, o protocolo de RT-*nested* PCR inicialmente utilizado para a amplificação das duas regiões acima referidas foi adaptado para a amplificação da região da IN. Os *primers* externos polO_for, polOc_for (directos) e HIV-vif2_rev (inverso) (**Tabela 2.1**), utilizados no passo inicial de retrotranscrição/primeira reacção de PCR, vão dar origem a dois fragmentos de DNA, de 1206 e 1410 pb, e, relativamente à segunda reacção de PCR, foi utilizado o par de *primers* internos polI_for/HIV-pol8_rev (**Tabela 2.1**). O *primer* directo polI_for hibrida na região codificante da RNase H da RT, enquanto que o *primer* inverso HIV-pol8_rev vai hibridar na região codificante da integrase, com a sua extremidade 5' apenas a trinta e dois nucleótidos do final do gene *pol*, permitindo a amplificação de um fragmento de 906 pb (**Figura 2.2**).

2.6.1.1 Retrotranscrição e primeira reacção de amplificação

Para a transcrição reversa do RNA viral e a primeira reacção de amplificação foram planeadas duas misturas de reacção em microtubos livres de RNases e DNases de 0,5 ml (*Multiply[®]-Pro*, Sarstedt, Alemanha). A mistura de reacção 1 foi preparada pela adição de 7 µl de água livre de RNases (Analytik Jena AG, Alemanha), 1 µl de uma mistura de dNTPs (concentração final, 0,2 mM cada), 1 µl de cada um dos *primers* externos para a região da RT/PR ou IN (**Tabela 2.1**) (concentração final, 0,2 µM cada), 2,5 µl de uma solução de DTT (concentração final 5 mM) e 0,5 µl de inibidor de RNases RNasin[®] (concentração final 0,4 U/µl) (Promega, EUA). A mistura de reacção 2 foi preparada com 14 µl de água livre de RNases (Analytik Jena AG, Alemanha), 10 µl de tampão de RT-PCR 5X (concentração final de 1,5 mM de MgCl₂) e 1 µl de mistura de enzimas (RT de AMV, polimerases de DNA Taq e Tgo, concentração final 0,02 U/µl), perfazendo um volume final de 25 µl por cada mistura de reacção. A preparação das misturas de reacção foi realizada em câmara de fluxo laminar (*Laminar Flow Cabinet-class 100*, Flow Laboratories, EUA), em “sala limpa”, na ausência de qualquer extracto de ácido nucleico. Em seguida, num compartimento físico separado, foram adicionados 10 µl de solução aquosa de RNA (utilizando pontas de micropipeta com filtro *ART 10 REACHTM*, Molecular BioProducts, EUA) à mistura de reacção 1, tendo-se posteriormente acrescentado os restantes 25 µl da mistura de reacção 2, perfazendo um volume final de 50 µl. As misturas de reacção foram sempre mantidas no gelo, durante todo o procedimento experimental. Após o término da primeira reacção

de amplificação, 2 µl do produto amplificado (cDNA) foram transferidos para novas misturas de reacção, como matriz de uma segunda amplificação por PCR.

2.6.1.2 Segunda reacção de amplificação

Na segunda reacção de amplificação foi utilizado o sistema *illustra™ puReTaq Ready-To-Go PCR Beads* (GE Healthcare, Inglaterra), no qual cada tubo de PCR (0,5 ml) contém uma pequena esfera liofilizada, otimizada especificamente para um volume final de reacção de 25 µl. Estas esferas contêm na sua composição estabilizadores, BSA (do inglês, *bovine serum albumin*), dNTPs, aproximadamente 2,5 unidades de polimerase de DNA *puReTaq* e os componentes de tampão de reacção. A reconstituição de cada mistura de PCR foi feita com uma solução aquosa ultrapura *Milli-Q®* (*Reagent Grade Ultrapure Systems*, Millipore, EUA) contendo cada um dos *primers* internos para as respectivas regiões em estudo (PR3/PR4, RT1/RT4 ou polI_for/HIV-pol8_rev), para uma concentração final de 0,6 µM de cada *primer*, 0,2 mM de cada dNTP, 50 mM de KCl e 1,5 mM de MgCl₂ em 10 mM de Tris-HCl, pH 9,0. É de referir que em todas as experiências realizadas de transcrição reversa *in vitro* e de amplificação foi sempre adicionado um controlo de contaminação (substituição do RNA/DNA matriz por água).

2.6.1.3 Condições de retrotranscrição *in vitro*/amplificação por *nested* PCR

As reacções de amplificação por RT-*nested* PCR foram realizadas no termociclador *iCycler* (Bio-Rad, EUA) e as condições experimentais utilizadas encontram-se descritas nas **Tabelas 2.2** e **2.3**. O passo inicial de retrotranscrição do RNA viral ocorreu a 45°C, durante 30 minutos. Os programas de amplificação realizados consistiram numa fase de desnaturação inicial, seguida de vários ciclos de amplificação (com desnaturação, hibridação dos *primers* e polimerização) e um período de extensão/polimerização final (condições nas **Tabelas 2.2** e **2.3**). Os produtos de amplificação eram mantidos a 4°C até serem analisados.

Tabela 2.2. Condições de retrotranscrição *in vitro* e de amplificação (1ª reacção de PCR) do RNA viral, para as regiões da protease, transcriptase reversa e integrase do gene *pol* do HIV-1.

Etapas e Condições da Retrotranscrição/1ª Reacção de PCR	
Retrotranscrição	45°C, 30 min
Desnaturação inicial [#]	94°C, 2 min
Amplificação	94°C, 30 seg; 45°C*, 30 seg; 68°C, 1 min, 10X
	94°C, 30 seg; 50°C*, 30 seg; 68°C, 2 min, 10X
	94°C, 30 seg; 55°C*, 30 seg; 68°C, 3 min, 15X
Extensão final	68°C, 7 min
Manutenção	4°C, ∞

*Na primeira reacção de amplificação correspondente à região da IN a temperatura de hibridação dos *primers* foi alterada para 56°C, sendo a fase de polimerização realizada durante 30 segundos e o número de ciclos de amplificação de 35.

[#]A desnaturação inicial tem como objectivo inactivar a AMV.

Tabela 2.3. Condições de amplificação (2ª reacção de PCR) do cDNA matriz, para as regiões da protease, transcriptase reversa e integrase do gene *pol* do HIV-1.

Etapas e Condições da 2ª Reacção de PCR	
Desnaturação inicial	94°C, 5 min
Amplificação*	94°C, 30 seg; 50°C, 30 seg; 72°C, 1 min, 35X
Extensão final	72°C, 7 min
Manutenção	4°C, ∞

*Na segunda reacção de amplificação correspondente à região da IN, a temperatura de hibridação dos *primers* foi de 62°C, o tempo de polimerização de 40 segundos e o número de ciclos de amplificação de 32.

2.6.2 Retrotranscrição com *primers* hexaméricos aleatórios

Nos casos em que não foi possível obter um resultado positivo de amplificação por RT-nested PCR, para qualquer uma das regiões em estudo (PR, RT ou IN), utilizando o protocolo descrito anteriormente, foi decidido a execução de uma metodologia alternativa baseada num passo inicial de retrotranscrição *in vitro* com *primers*

hexaméricos aleatórios, seguido da amplificação do cDNA por *nested* PCR convencional (dois passos separados e consecutivos). A reacção de retrotranscrição foi realizada recorrendo ao sistema *RevertAid™ H Minus First Strand cDNA Synthesis* (Fermentas, Canadá), segundo as especificações do fabricante, que se descrevem, brevemente, de seguida. Em câmara de fluxo laminar (*Laminar Flow Cabinet-class 100*, Flow Laboratories, EUA), em “sala limpa”, foram preparadas duas misturas de reacção, em microtubos livres de RNases e DNases de 0,5 ml (*Multiply®-Pro*, Sarstedt, Alemanha). Para a mistura de reacção 1 foram adicionados 4 µl de tampão de reacção 5X (250 mM Tris-HCl, pH 8,3, 250 mM KCl, 20 mM MgCl₂, 50 mM DTT), 1 µl do inibidor de RNases *RiboLock™*, 2 µl de uma mistura de dNTPs a 10 mM (concentração final 1 mM) e 1 µl da enzima *RevertAid™ H Minus M-MuLV Reverse Transcriptase* (concentração final 10 U/µl). Para a mistura de reacção 2 foram misturados 6 µl de água tratada com DEPC (do inglês, *diethylpyrocarbonate*) e 1 µl de solução de *primers* hexaméricos aleatórios (0,2 µg). Posteriormente, num compartimento físico separado, foram adicionados 5 µl de solução de RNA matriz (utilizando pontas de micropipeta com filtro *ART 10 REACH™*, Molecular BioProducts, EUA) à mistura de reacção 2, a qual foi colocada num banho a 70°C, durante 5 minutos e logo de seguida posta em gelo, durante 5 minutos. Terminado este passo, adicionou-se ao microtubo anterior a mistura de reacção 1, perfazendo um volume final de 20 µl/reacção. De seguida, foi iniciada a etapa de síntese do cDNA no termociclador *iCycler* (Bio-Rad, EUA). O programa escolhido consistiu num passo inicial de hibridação dos *primers* hexaméricos aleatórios à cadeia simples de RNA, seguindo-se a conversão do RNA em cDNA e finalmente a paragem da reacção de retrotranscrição (**Tabela 2.4**). O cDNA foi armazenado a -80°C, até ser utilizado para a amplificação por *nested* PCR.

Tabela 2.4. Condições de retrotranscrição *in vitro* do RNA viral com *primers* hexaméricos aleatórios.

Hibridação dos <i>primers</i>	25°C, 5 min
Retrotranscrição	42°C, 60 min
Inactivação da RT	70°C, 5 min
Manutenção	4°C, ∞

2.6.2.1 Amplificação do cDNA por *nested* PCR

Na primeira e segunda reacções de amplificação foi utilizado o sistema *illustra™ puReTaq Ready-To-Go PCR Beads* (GE Healthcare, Inglaterra), segundo as indicações do fabricante, utilizando as especificações descritas anteriormente. Como matriz para as duas reacções de PCR foram usados, respectivamente, 3 µl de solução de cDNA e de produto amplificado da 1ª reacção. A amplificação foi realizada nas condições experimentais já referidas (**Tabelas 2.2 e 2.3**), anulando o passo inicial de retrotranscrição e alterando a temperatura de polimerização de 68°C para 72°C (**Tabela 2.2**).

2.7 Observação dos Fragmentos de DNA Amplificados por RT-Nested PCR

Para observação dos produtos de PCR obtidos, após separação por electroforese em gel de agarose (Agarose MP, Boehringer, Alemanha), a 1,2% (p/v), em tampão TAE (Tris-Acetato-EDTA) 0,5X (**Anexo 2**), com brometo de etídeo (0,5 µg/ml), foi utilizado o sistema de mini-gel *GNA100* (Pharmacia Biotech, Suécia) e a fonte de corrente eléctrica *Power Pac 300* (Bio-Rad, EUA). Após 30 minutos de polimerização de uma solução de 50-100 ml de gel, seguiu-se a aplicação de 5 µl de cada amostra, misturados com cerca de 1 µl de tampão de aplicação (**Anexo 2**). A voltagem aplicada rondava os 100-110 V. O tamanho dos fragmentos obtidos foi estimado por comparação com o padrão de migração do marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Canadá). Os fragmentos de DNA foram visualizados após exposição do gel a radiação ultravioleta ($\lambda=254$ nm), sendo a imagem captada e digitalizada num aparelho *Gel-Doc XR* (Bio-Rad, EUA).

2.8 Purificação dos Fragmentos de DNA Amplificados

A partir do produto de PCR

Após observação no gel de agarose (Agarose MP, Boehringer, Alemanha), a presença de uma banda específica, de tamanho esperado, e de quantidade razoável de DNA [concentração estimada por comparação com o marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Canadá)], eram as premissas necessárias para avançar com a purificação dos fragmentos de DNA amplificados directamente a partir do produto de PCR. Para tal, recorreu-se ao sistema *QIAquick PCR Purification Kit* (Qiagen, Alemanha), utilizado de acordo com as especificações do

fabricante. Resumidamente, a 15 µl do produto de PCR eram adicionados 75 µl de tampão PB (proporção 1:5). Logo de seguida, esta mistura era aplicada numa *QIAquick spin column*, previamente adaptada a um microtubo de 2 ml, seguindo-se um passo de centrifugação a 13000 rpm, durante 1 minuto (*Centrifuge 5415-D*, Eppendorf, Alemanha). Após a remoção do eluato, seguia-se um passo de lavagem, por adição de 750 µl de tampão PE à *QIAquick spin column* e dupla centrifugação, nas condições anteriores. Finalmente, o DNA era eluído, por centrifugação a 13000 rpm, durante 1 minuto (*Centrifuge 5415-D*, Eppendorf, Alemanha), com 35 µl de água ultrapura *Milli-Q®* (*Reagent Grade Ultrapure Systems*, Millipore, EUA). O sucesso deste passo de purificação era confirmado por observação de uma alíquota da solução de DNA purificado após electroforese em gel de agarose (Agarose MP, Boehringer, Alemanha), nas condições descritas na secção anterior.

A partir do gel de agarose

Em duas amostras nas quais se observou, após migração electroforética, a presença de produtos de amplificação inespecíficos, procedeu-se à purificação dos fragmentos de DNA amplificados, de tamanho pretendido, a partir do gel de agarose. Este procedimento foi realizado recorrendo ao sistema *innuPREP Gel Extraction Kit* (Analytik Jena AG, Alemanha), segundo as instruções do fabricante. Resumidamente, após identificação do fragmento de DNA pretendido no gel de agarose (Agarose MP, Boehringer, Alemanha), sob radiação ultravioleta, este foi cuidadosamente removido, com recurso a um bisturi estéril, e colocado num microtubo de 2 ml. Para a dissolução completa da agarose, adicionou-se uma solução de solubilização³ e colocou-se num banho-maria a 50°C, durante 10 minutos. Em seguida, foi realizado o passo de ligação do DNA às colunas (*spin filters*) e posterior eluição por centrifugação a 12000 rpm, durante 1 minuto (*Centrifuge 5415-D*, Eppendorf, Alemanha). Procedeu-se de imediato a duas lavagens consecutivas do DNA com 500 e 750 µl das soluções de lavagem HS (do inglês, *high salt*) e LS (do inglês, *low salt*), respectivamente (12000 rpm, durante 1 minuto). Finalmente, o produto purificado foi eluído com o tampão de eluição para um volume final de 35 µl. Tal como descrito na secção anterior, uma alíquota da solução de

³ Para fragmentos de agarose com peso superior a 150 mg, adiciona-se 1 ml de solução de solubilização; para peso inferior, adiciona-se 500 µl da mesma solução.

DNA purificado foi analisada por electroforese em gel de agarose (Agarose MP, Boehringer, Alemanha).

2.9 Clonagem dos Produtos de PCR no Vector Plasmídico *pGEM[®]-T Easy*

Após observação dos produtos de PCR por electroforese em gel de agarose (Agarose MP, Boehringer, Alemanha), verificou-se que alguns dos fragmentos de DNA amplificados apresentavam-se em quantidades reduzidas ou vinham acompanhados de produtos amplificados inespecíficos. Nestes casos, recorria-se ao processo de clonagem, previamente à sequenciação, sendo utilizado o sistema *pGEM[®]-T Easy Vector* (Promega, EUA), segundo as especificações do fabricante.

2.9.1 Caracterização do vector plasmídico *pGEM[®]-T Easy*

O vector plasmídico *pGEM[®]-T Easy* (Promega, EUA), é fornecido linearizado e

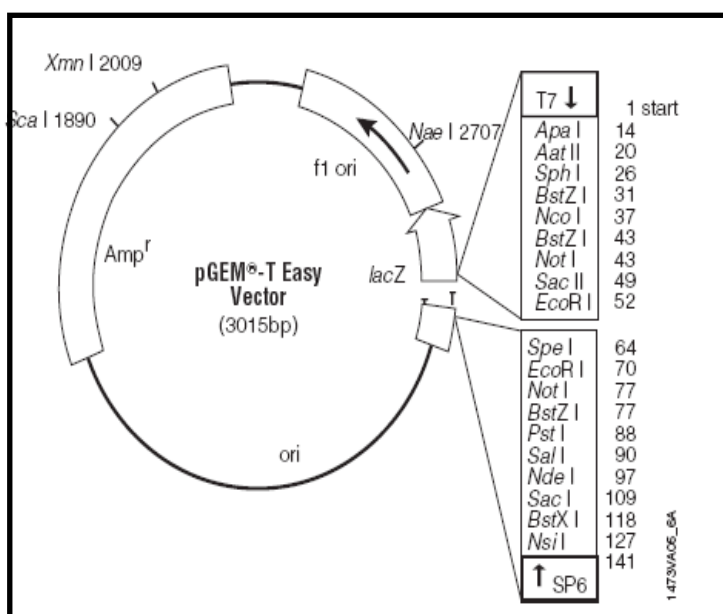


Figura 2.3. Mapa físico do vector plasmídico de 3015 pb, *pGEM-T[®] Easy* (Promega, EUA). Na figura estão representadas os locais de reconhecimento de diversas enzimas de restrição na região de clonagem múltipla e respectivas coordenadas. A origem de replicação está indicada por ori, o gene *bla*, que codifica para a resistência à ampicilina, está representado por *Amp^r* e *lacZ* representa a sequência codificante do fragmento α da β -galactosidase (adaptado de www.promega.com).

apresenta resíduos de 3'-timidina, em ambas as extremidades, permitindo a fácil ligação de fragmentos de DNA amplificados pela polimerase de DNA Taq (esta adiciona um resíduo de adenosina extra às extremidades 3', no final de cada ciclo de amplificação). É um plasmídeo que contém os promotores de transcrição T7 e SP6, a flanquear a região de clonagem múltipla (Figura 2.3), a qual, por sua vez, se

encontra incluída no gene do α -péptido da enzima β -galactosidase (*lacZ*), o que permite a selecção de recombinantes por α -complementação. A inserção de um fragmento de

DNA nesta região impossibilita a expressão da forma activa da enzima, permitindo a triagem de clones bacterianos contendo moléculas de DNA plasmídico recombinante, que surgem como colónias brancas em meio contendo IPTG (isopropil- β -D-tiogalactopiranósido) e X-gal (5-bromo-4-cloro-3-indolil- β -D-galactopiranósido) (**Anexo 2**).

2.9.2 Etapas do processo de clonagem

O processo de clonagem é caracterizado por várias etapas, nomeadamente:

Ligação do fragmento de DNA ao vector plasmídico *pGEM[®]-T Easy*

A ligação do fragmento do DNA ao vector plasmídico foi realizada através da preparação de uma mistura de reacção constituída por 3 μ l do produto de PCR, 1 μ l de solução de vector plasmídico *pGEM[®]-T Easy* (Promega, EUA) (50 ng/ μ l), 5 μ l do tampão de ligação da ligase de DNA do fago T4 2X (**Anexo 2**) e 1 μ l da respectiva enzima (3 U/ μ l), para um volume final de 10 μ l, tendo-se posteriormente mantido a mistura de ligação a 4°C, em banho refrigerado (*FRIGITERM*, P-Selecta, Espanha), durante a noite, para obtenção de máximo rendimento de ligação.

Preparação de células competentes e sua transformação por choque térmico

Para as experiências de transformação, foi utilizada a estirpe JM109 de *Escherichia coli*, estirpe sobreprodutora do repressor Lac e naturalmente sensível à ampicilina. A estirpe JM109, congelada previamente a -80°C, em meio de criopreservação (**Anexo 2**), foi inoculada em 3 ml de meio LB líquido (**Anexo 2**) e colocada, durante a noite, a 37°C, com agitação (220 rpm). No dia seguinte, adicionou-se 200 μ l da suspensão celular a 20 ml do mesmo meio, controlando-se o crescimento bacteriano por medição de D.O.₆₀₀ (densidade óptica a 600 nm), até se atingir um valor compreendido entre 0,3-0,5. As células foram posteriormente recolhidas por centrifugação a 6000 rpm, durante 10 minutos, a 4°C (*Centrifuge J-21*, Beckman, EUA) e, logo de seguida, desprezou-se o sobrenadante e ressuspendeu-se em 1 ml de meio TSS frio (do inglês, *transformation and storage solution*) (**Anexo 2**). Para a transformação propriamente dita, foram adicionados 5 μ l de mistura de ligação previamente preparada a 100 μ l da suspensão de células competentes de *Escherichia coli* JM109 em meio TSS. O mesmo procedimento foi efectuado para os devidos controlos sendo adicionado, em substituição da mistura de ligação, 1 μ l de pKS

(*pBluescript* KS II, SIGMA, Alemanha), ao controlo positivo, e 5 µl de dH₂O, ao controlo negativo. Após homogeneização suave, as células foram colocadas no gelo durante 60 minutos, sendo, posteriormente, submetidas a choque térmico (banho-maria a 42°C, durante 90 segundos, e colocação imediata em gelo). Depois da adição de 900 µl de meio LB líquido, os microtubos foram incubados a 37°C, a 80 rpm, durante 1 hora. Terminada a incubação, seguiu-se um passo de centrifugação a 6000 rpm, durante 10 minutos (*Centrifuge* 5415-D, Eppendorf, Alemanha). O sobrenadante foi removido e o sedimento celular cuidadosamente ressuscitado em 200 µl do mesmo meio, os quais foram distribuídos por espalhamento (em condições de assepsia), em duas placas de Petri com meio LB agar (15 ml/placa), suplementado com ampicilina, X-gal e IPTG (**Anexo 2**). As placas foram incubadas a 37°C, durante a noite.

Seleção de clones bacterianos recombinantes

A adição de suplementos ao meio LB agar teve como propósito a selecção de clones bacterianos transformados (selecção com ampicilina) recombinantes (triagem branco/azul), os quais originam colónias de cor branca. Após o período de incubação, procedeu-se à repicagem das colónias brancas (em condições de assepsia), com palito estéril, inoculando-as em 3 ml de meio LB líquido com ampicilina (100 µg/ml) e incubando a 37°C, a 220 rpm, durante a noite, de modo a iniciar o protocolo de extracção do DNA plasmídico.

2.10 Extracção do DNA Plasmídico Pelo Método de Lise Alcalina

A extracção do DNA plasmídico foi realizada pelo método de lise alcalina (14). Para tal, e após transferência de 1,5 ml da suspensão bacteriana para microtubo, procedeu-se a um passo de centrifugação a 13500 rpm, durante 2 minutos (*Centrifuge* 5415-D, Eppendorf, Alemanha), tendo-se decantado o sobrenadante e vortexado brevemente o sedimento celular, a fim de facilitar o passo seguinte de ressuspensão em 100 µl de tampão TEG gelado (**Anexo 2**). Posteriormente, adicionou-se o mesmo volume de SDS alcalino (**Anexo 2**), preparado no momento, seguindo-se a mistura por inversão, para facilitar a lise celular. De seguida, foram adicionados 200 µl de acetato de potássio 3M (pH 4,8) frio, sendo o conteúdo dos tubos misturado por inversões sucessivas. Os restos celulares foram sedimentados por centrifugação a 13500 rpm, durante 15 minutos, a 4°C (*Centrifuge* 3k18, Sigma, Alemanha). O sobrenadante foi

transferido para novos microtubos, aos quais se adicionaram 600 µl de isopropanol (Merck, Alemanha), fazendo-se uma homogeneização por inversão, ao que se seguiu um novo passo de centrifugação, nas condições anteriores, mas à temperatura ambiente. O sobrenadante foi cuidadosamente decantado e o sedimento lavado com 250 µl de etanol (Merck, Alemanha) a 70% (v/v), previamente mantido a -20°C. Após centrifugação, nas mesmas condições, durante 5 minutos, o sedimento de DNA plasmídico foi seco num evaporador (*Concentrator 5301*, Eppendorf, Alemanha). Finalmente, o sedimento seco foi ressuscitado em 30 µl de água ultrapura *Milli-Q*[®] (*Reagent Grade Ultrapure Systems*, Millipore, EUA), ao qual se adicionou RNase A (concentração final de 50 µg/ml), ao que se seguiu uma incubação em banho-maria a 37°C, durante 30 minutos, a fim de permitir a degradação do RNA. Se não utilizado de imediato, o DNA plasmídico foi mantido a -20°C.

2.10.1 Observação das moléculas de DNA plasmídico

Para confirmação da clonagem, foi efectuada uma electroforese em gel de agarose (Agarose MP, Boehringer, Alemanha), a 1% (p/v), em tampão TAE 0,5X (**Anexo 2**), corado com brometo de etídeo (0,5 µg/ml). A selecção das moléculas de DNA plasmídico potencialmente recombinantes foi baseada na comparação dos padrões de migração apresentados pelos diferentes plasmídeos extraídos. As moléculas que apresentavam uma migração mais lenta no gel eram seleccionadas para hidrólise com a enzima de restrição EcoRI, visto que esta permite a excisão do fragmento de DNA clonado no vector plasmídico *pGEM*[®]-*T Easy* (Promega, EUA) (**Figura 2.3**).

2.11 Hidrólise do DNA Plasmídico com a Enzima de Restrição EcoRI

A presença do fragmento clonado foi confirmada por hidrólise dos plasmídeos potencialmente recombinantes com a endonuclease de restrição EcoRI (Fermentas, Canadá). Esta enzima reconhece dois locais presentes no sítio de clonagem múltipla, flaqueando o local de inserção do fragmento de DNA heterólogo (**Figura 2.3**). A digestão dos DNAs plasmídicos seleccionados foi realizada num volume final de 15 µl, contendo 10 µl de solução de DNA extraído, 1 µl de EcoRI (10 U/µl), 1,5 µl de tampão EcoRI 10X (**Anexo 2**), perfazendo-se o restante volume (2,5 µl) com água desionizada ultrapura *Milli-Q*[®] (*Reagent Grade Ultrapure Systems*, Millipore, EUA). Os produtos de digestão foram analisados após migração electroforética, em gel de agarose (Agarose

MP, Boehringer, Alemanha), a 1,2% (p/v), em tampão TAE 0,5X (**Anexo 2**), corado com brometo de etídeo (0,5 µg/ml).

2.12 Extracção e Purificação do DNA Plasmídico para Sequenciação

A extracção e purificação do DNA plasmídico recombinante, para sequenciação nucleotídica, foram efectuadas a partir de 3,5 ml de cultura saturada, deixada a crescer durante a noite, em meio LB com ampicilina (100 µg/ml) (37°C, 220 rpm), recorrendo ao sistema *Wizard® Plus SV Minipreps DNA Purification System* (Promega, EUA), utilizado de acordo com as instruções do fabricante. Muito sucintamente, após a produção de um lisado celular clarificado, seguiu-se a ligação do DNA plasmídico às *spin columns* e dois passos consecutivos de lavagem. A eluição final do DNA plasmídico purificado foi realizada com água desionizada ultrapura *Milli-Q®* (*Reagent Grade Ultrapure Systems*, Millipore, EUA), livre de RNases e DNases, para um volume final de 100 µl. A confirmação da purificação foi feita por visualização do DNA plasmídico após migração por electroforese em gel de agarose (Agarose MP, Boehringer, Alemanha), a 1% (p/v), em tampão TAE 0,5X (**Anexo 2**), corado com brometo de etídeo (0,5 µg/ml).

2.13 Sequenciação do DNA

Depois da purificação das amostras de DNA (descrição na secção 2.8), as mesmas foram enviadas para sequenciação nucleotídica, ao cuidado da empresa STAB VIDA (Caparica, Portugal). O volume enviado de amostra, quando purificada directamente a partir do produto de PCR, foi de 15 µl (para um mínimo de 400 ng de DNA/reacção). Quanto aos *primers* (PR3/PR4, RT1/RT4 e polI_for/HIV-pol8_rev), foram enviados 50 pmoles/reacção de sequenciação, dado que se pretendia a obtenção da sequência nos dois sentidos. Relativamente aos fragmentos de DNA clonados, foram enviados 40 µl de solução de DNA plasmídico purificado (para um mínimo de 1 µg de DNA/reacção). Para a sequenciação dos fragmentos clonados da PR (460 pb) e da RT (650 pb) foram utilizados os *primers* universais T7 (directo) e SP6 (inverso), complementares às sequências que flanqueiam o local de clonagem múltipla do vector *pGEM®-T Easy* (Promega, EUA). Devido ao tamanho relativamente grande apresentado pelo fragmento clonado da IN (906 pb), a sequenciação destas amostras foi realizada com o par de *primers* internos, utilizados na segunda reacção de PCR, nomeadamente, polI_for e

HIV-pol8_rev. Em todos os casos, a sequenciação foi realizada pelo método de terminação de cadeia descrito por Sanger *et al* (1985), implementado num protocolo de sequenciação automática.

2.14 Análise Bioinformática de Sequências do HIV-1

2.14.1 Sequências nucleotídicas

2.14.1.1 Análise dos cromatogramas e edição das sequências

Os cromatogramas relativos às reacções de sequenciação foram cuidadosamente analisados e corrigidos, a fim de se proceder à edição de sequências nucleotídicas “consenso”, entre as sequências das cadeias directa e inversa. Para tal, recorreu-se ao programa *BioEdit Sequence Alignment Editor*, v.7.0.5.3. (87, disponível em <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>).

2.14.1.2 Pesquisa de homologia genética

A fim de confirmar a sua identidade, as sequências nucleotídicas editadas foram submetidas a uma pesquisa rápida de homologia genética, recorrendo-se ao programa *BLAST* (do inglês, *Basic Local Alignment Search Tool*) (disponível em <http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Este programa, para além de calcular a percentagem de semelhança entre as sequências, estima também a probabilidade da homologia genética encontrada ser devida ao acaso (*E value*, 0-1).

2.14.1.3 Determinação automática dos genótipos virais

Inicialmente, recorreu-se ao programa *NCBI HIV Genotyping Program* (disponível em <http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>), que pesquisa regiões de similaridade local entre a sequência em estudo e sequências de HIV de referência, de genótipo conhecido, existentes numa base de dados específica (183). Esta genotipagem automatizada apresenta, no entanto, a limitação de ser exclusivamente baseada na pesquisa de graus de similaridade genética.

Mais tarde, para confirmação dos resultados obtidos anteriormente, recorreu-se ao programa *REGA HIV-1 & 2 Automated Subtyping Tool* v.2.0 (disponível em <http://jose.med.kuleuven.be/genotypetool/html/subtypinghiv.html>). Esta ferramenta bioinformática permite a execução de uma análise filogenética automatizada rápida, combinada com métodos de *bootscanning* para a genotipagem de fragmentos subgenómicos localizados (50).

2.14.1.4 Construção de árvores filogenéticas

A genotipagem foi realizada com recurso à construção de árvores filogenéticas (regiões da PR, RT, IN e uma sequência concatenada), tendo-se alinhado as sequências nucleotídicas editadas recorrendo ao programa *ClustalW2* v.2.0 (218, disponível em www.ebi.ac.uk/clustalw/). Para cada uma das regiões, o alinhamento múltiplo foi realizado com as sequências em estudo, sequências de referência do HIV-1 do grupo M [subtipos, sub-subtipos e duas CRFs (02_AG e 14_BG)], facultadas na base de dados de *Los Alamos* (disponível em <http://hiv-web.lanl.gov/>), e uma sequência do grupo O, como *outgroup*. O alinhamento múltiplo obtido foi editado manualmente no programa *BioEdit* v.7.0.5.3. (87), tendo-se removido os espaços abertos. Foram ainda eliminados os segmentos nucleotídicos que não correspondiam às regiões PR, RT ou IN, propriamente ditas (embora pudessem ter sido amplificadas). Finalmente, a construção das árvores filogenéticas foi efectuada recorrendo ao programa *MEGA* (do inglês, *Molecular Evolutionary Genetics Analysis*) v.4.0 (214, disponível em <http://www.megasoftware.net/>). A sua construção baseou-se no método de *Neighbor-Joining* (184), utilizando o modelo de 2 parâmetros de Kimura (108) para o cálculo da matriz de distâncias genéticas (assumindo uma frequência igual de bases, uma taxa de transversões igual à taxa de transições e com a opção *pairwise deletion*). A robustez da inferência filogenética e, conseqüentemente, das topologias sugeridas foi avaliada através do método de *bootstrapping*, com base em 1000 amostragens aleatórias, considerando-se estatisticamente significativos valores iguais ou superiores a 70%.

2.14.1.5 Análise de sequências divergentes

Para a análise das sequências cujos ramos de suporte na árvore evidenciavam um posicionamento ambíguo, sugerindo divergência filogenética elevada, procedeu-se à:

Pesquisa de sequências de natureza recombinante, por métodos de *bootscanning*, com utilização do programa *SimPlot* v.3.5.1. (120, disponível em <http://sray.med.som.jhmi.edu/SCRsoftware/simplot/>);

Pesquisa de fenómenos de hipermutação por ocorrência de hiperadenilação G→A dependente da APOBEC, recorrendo ao programa *Hypermut* 2.0 (181, disponível em http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html?sample_input=1).

2.14.1.6 Análise das distâncias genéticas

O cálculo das distâncias genéticas entre grupos de sequências seleccionados foi realizado recorrendo-se ao programa *Subtyping Distance Tool* (SUDI, disponível em <http://www.hiv.lanl.gov/content/sequence/SUDI/sudi.html>).

2.14.2 Sequências de aminoácidos

A análise de mutações associadas à diminuição da susceptibilidade aos anti-retrovirais inibidores enzimáticos foi efectuada recorrendo-se ao programa *HIVdb*, implementado através do *Genotypic Resistance Interpretation Algorithm* da *Stanford University HIV Drug Resistance Database* (disponível em http://hivdb.stanford.edu/pages/algs/sierra_sequence.html). De realçar que as sequências estudadas foram comparadas com uma sequência de aminoácidos consenso do subtipo B. Um relatório interpretativo é obtido para cada sequência, identificando as mutações associadas a resistência como principais (*major*) ou acessórias (*minor*), realçando ainda a existência de outros polimorfismos genéticos não associados a resistência aos anti-retrovirais (*other mutations*). Com base em diversos tipos de estudos, o algoritmo atribui ainda uma pontuação a cada tipo de mutação encontrada, relativamente a cada um dos anti-retrovirais licenciados, reflectindo o grau de probabilidade dessa mutação em particular estar associada à aquisição de resistência fenotípica a esse mesmo fármaco. Para uma dada sequência aminoacídica, a estimativa de resistência a cada anti-retroviral é então obtida pela soma das pontuações particulares atribuídas a cada uma das mutações encontradas. Posteriormente, a sua expressão fenotípica é classificada em cinco categorias diferentes (205):

Susceptibilidade (0-9 pontos) - os vírus apresentam um fenótipo de susceptibilidade total, comparativamente com a estirpe selvagem;

Resistência potencial de nível baixo (10-14 pontos) - indica que os vírus não apresentam uma susceptibilidade diminuída; contudo, possuem mutações indicativas de eventual exposição prévia a uma determinada classe terapêutica anti-retroviral (estas mutações, por si só, não causam resistência aos anti-retrovirais);

Resistência de nível baixo (15-29 pontos) - os vírus apresentam susceptibilidade reduzida ao fármaco *in vitro* e/ou os indivíduos infectados com estes vírus podem

apresentar uma resposta virológica sub-ótima ao tratamento, comparativamente com os infectados com vírus selvagem;

Resistência de nível intermédio (30-59 pontos) - o genótipo viral sugere um grau de resistência ao fármaco com um impacto significativo na sua eficácia;

Resistência de nível elevado (superior a 60 pontos) - traduz uma quase ausência de resposta terapêutica *in vivo* e/ou que vírus com o genótipo similar apresentaram os níveis mais elevados de resistência *in vitro*.

2.15 Análise Estatística

Aos dados epidemiológicos que caracterizam a população estudada, de natureza categórica, como o sexo e origem geográfica dos indivíduos e respectivos parceiros sexuais, terapêutica anti-retroviral, partilha de seringas, relações sexuais desprotegidas, infecção recente ou antiga e estágio da infecção, aplicou-se uma análise de distribuição. Relativamente às variáveis numéricas, como idade dos indivíduos e carga viral, para além da análise de distribuição, calculou-se ainda a média e o desvio padrão. A associação de polimorfismos genéticos específicos (e respectivo perfil fenotípico) com os diferentes subtipos virais em circulação na população estudada foi verificada recorrendo-se ao teste exacto de Fisher bilateral, aplicado à análise de tabelas de contingência 2x2 (disponível em <http://www.physics.csbsju.edu/stats/>), considerando um valor de significância de 5% (22).

3.Resultados

3.1. Caracterização Epidemiológica da População Estudada

As variáveis demográficas, comportamentais, clínicas e laboratoriais disponíveis do total de 51 indivíduos em estudo, todos utilizadores de drogas por via endovenosa e seropositivos para HIV-1, foram analisadas e os resultados desta análise de distribuição encontram-se representados na **Figura 3.1**.

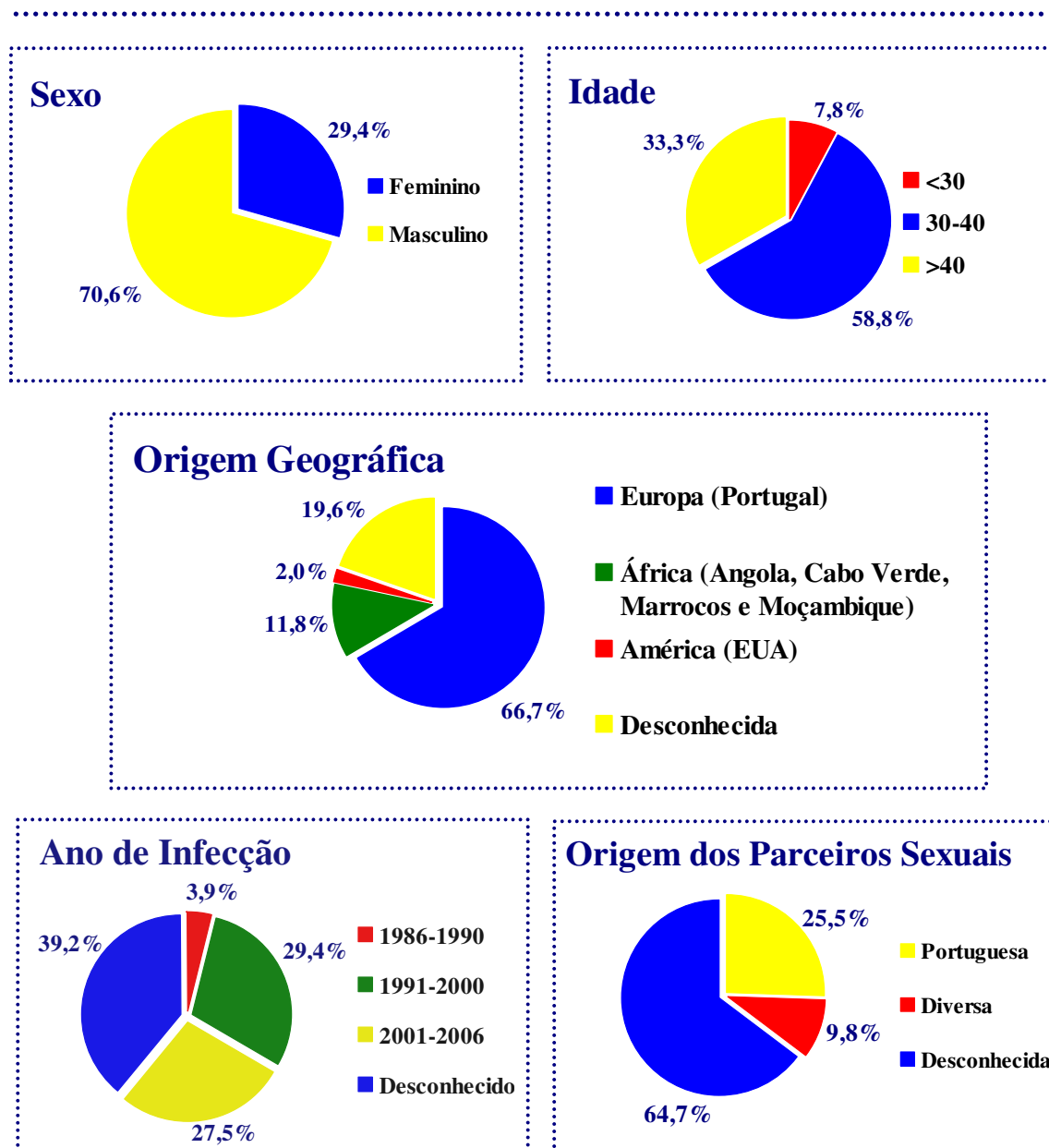


Figura 3.1. Representação gráfica dos resultados (em percentagem) da análise de distribuição relativa às seguintes variáveis: sexo, idade e origem geográfica dos indivíduos, ano de infecção, origem dos parceiros sexuais, estágio da infecção, terapêutica anti-retroviral, carga viral, prática de partilha de seringas e de relações sexuais não protegidas.

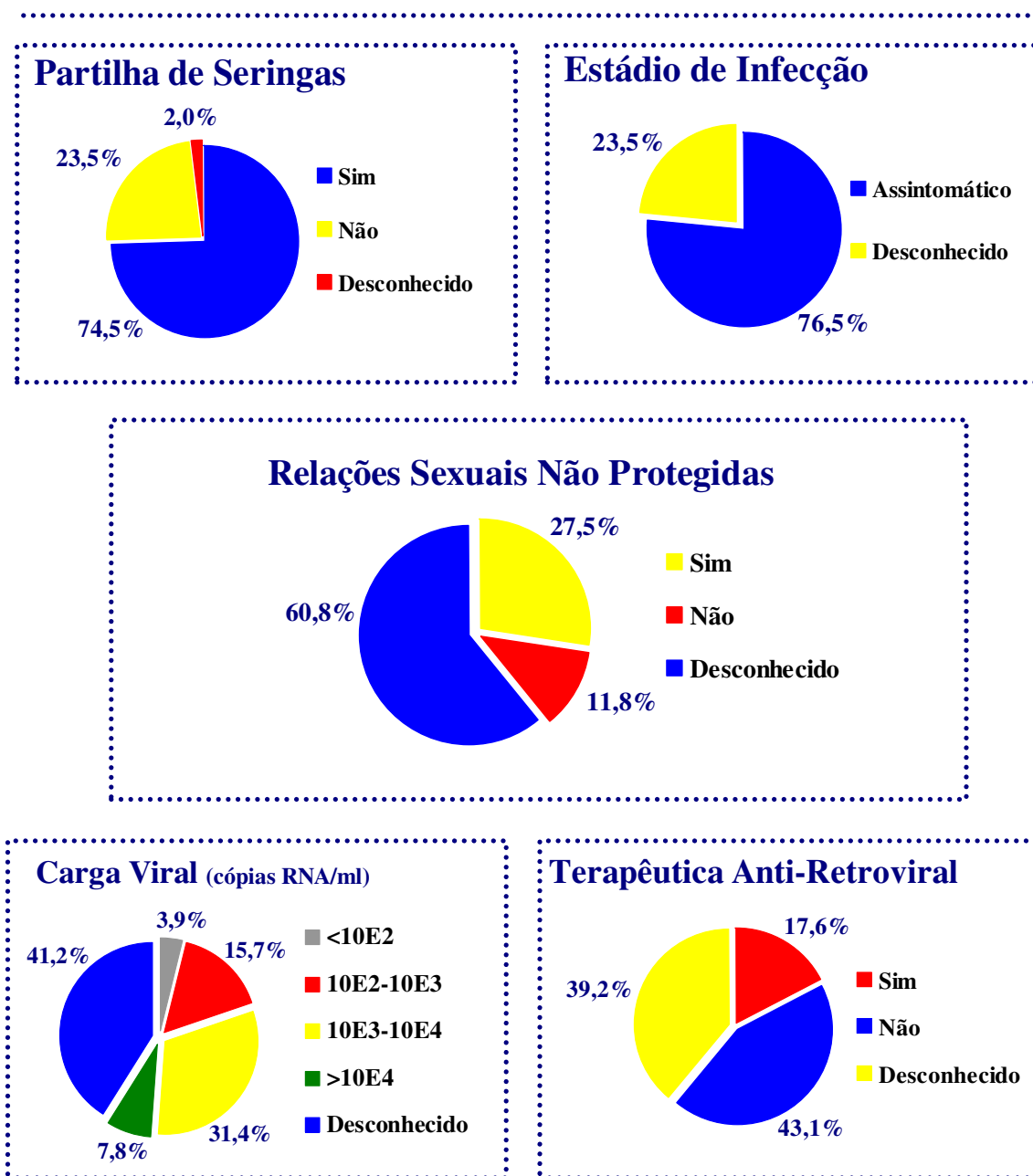


Figura 3.1. Continuação

Como se pode observar na **Figura 3.1**, o conjunto das 51 amostras provém maioritariamente de indivíduos de nacionalidade portuguesa (66,7%), do género masculino (70,6%) e com idade compreendida entre os 30 e os 40 anos (idade média de $37,5 \pm 5,68$ anos). A maioria das infecções (29,4%) foi adquirida entre 1991 e 2000 (informação desconhecida para 39,2% da amostra). No que respeita aos comportamentos de risco, a partilha de material de injeção foi relatada por 74,5% dos

indivíduos, enquanto que a prática de relações sexuais não protegidas foi afirmada por 27,5% (embora para esta variável exista um número considerável de indivíduos cuja resposta é desconhecida). A existência de parceiros sexuais com origens geográficas diversas (para além da exclusivamente portuguesa) foi relatada por 9,8% da amostra. Quanto ao estágio de infecção, todos os indivíduos para os quais esse dado era conhecido afirmaram encontrar-se em fase assintomática (76,5% da amostra). Relativamente à terapêutica anti-retroviral, os dados disponíveis reportam-se somente a cerca de 60,7% da amostra (à data da colheita, 17,6% dos indivíduos encontrava-se sob qualquer tipo de terapêutica anti-retroviral, enquanto que 43,1% não). Finalmente, no que respeita aos níveis de RNA de HIV-1 no plasma, estes dados foram reportados somente para as amostras provenientes do INSA (n=30), variando os valores entre 470 e $1,9 \times 10^6$ cópias de RNA/ml (valor médio de $9,7 \times 10^4 \pm 3,44 \times 10^5$).

3.2.Caracterização dos Fragmentos de DNA a Amplificar

A origem e preparação dos *primers* utilizados para a amplificação, por RT-nested PCR, das regiões codificantes da transcriptase reversa (RT), da protease (PR) e da integrase (IN) do gene *pol* do HIV-1 encontram-se descritas no capítulo anterior (2.5. Origem e Preparação dos *Primers*), estando as respectivas sequências e coordenadas indicadas na **Tabela 2.1**. A numeração dos nucleótidos foi atribuída com base na sequência completa da estirpe de referência HXB2 de HIV-1 (disponível em <http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2009/hiv1dna.pdf>), o que permitiu uma caracterização prévia dos fragmentos de DNA que se pretendiam amplificar (**Tabela 3.1**).

Tabela 3.1. Caracterização breve dos fragmentos do gene *pol* do HIV-1 a amplificar por RT-nested PCR.

Proteína codificada (gene <i>pol</i>)	Localização da região amplificada (em HXB2)	Tamanho do fragmento (pb)
Protease	De 13 nucleótidos a jusante do início da região codificante da proteína p6 (gene <i>gag</i>) até 55 nucleótidos a jusante do início da região codificante da subunidade p51 da RT (gene <i>pol</i>)	460
Transcritase reversa	A totalidade da região amplificada encontra-se no interior da região codificante da subunidade p51 da RT	650
Integrase	De 72 nucleótidos a montante do final da região codificante do motivo RNase H da RT até 32 nucleótidos a montante do final do gene <i>pol</i>	906

3.3. Amplificação por RT-Nested PCR

Os protocolos de RT-nested PCR utilizados para a amplificação dos fragmentos da PR e da RT não necessitaram de um período inicial de otimização experimental, visto que as condições de amplificação (referidas no subcapítulo 2.6.1.3) já tinham sido otimizadas em estudos anteriores (207). Para a região da IN, no entanto, foi necessário fazer uma adaptação do procedimento experimental descrito por Braga (15), somente no que respeita às condições de retrotranscrição e amplificação (descrito em 2.6.1.3).

Como referido no capítulo 2, foi realizada a extracção do RNA viral para um total de 51 amostras de plasma de indivíduos seropositivos para HIV-1, tendo sido este posteriormente utilizado como matriz para a realização da etapa da retrotranscrição *in vitro*, seguida de amplificação do cDNA por *nested* PCR. Tal como descrito anteriormente, foram utilizadas, em sucessão, duas metodologias distintas de retrotranscrição *in vitro* do RNA viral.

3.3.1 Retrotranscrição *in vitro* com *primers* específicos

Os resultados de amplificação obtidos por *nested* PCR, após retrotranscrição *in vitro* com *primers* específicos [sistema *Titan One Tube RT-PCR* (Roche Applied Science, Alemanha)], encontram-se representados na **Figura 3.2**. Foi possível obter um total de 40 (78,4% de positividade), 41 (80,4%) e 39 (76,5%) amplicões, respectivamente, para PR, RT e IN.

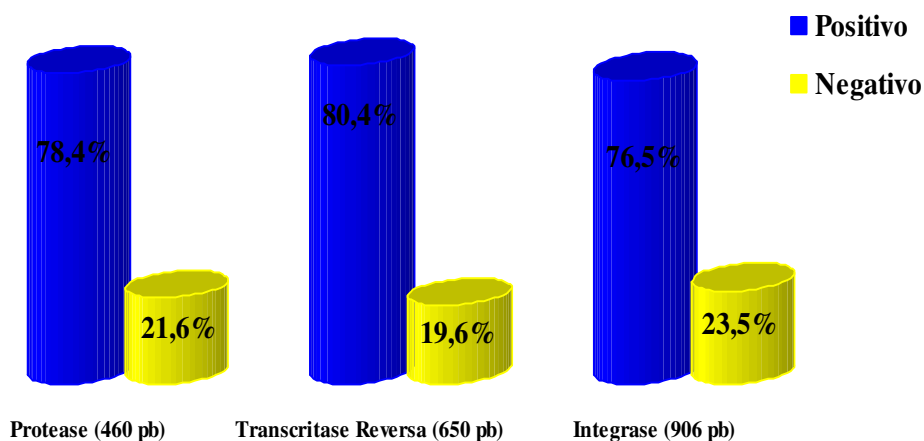


Figura 3.2. Representação gráfica dos resultados obtidos por RT-nested PCR (retrotranscrição *in vitro* com *primers* específicos) para as três regiões do gene *pol* estudadas (n=51 amostras).

A análise dos produtos amplificados por electroforese em gel de agarose revelou a presença de fragmentos específicos, com o tamanho esperado de 460, 650 e 906 pb, respectivamente para PR, RT e IN (a título de exemplo, na **Figura 3.3**, amostras identificadas nos poços 1-5, para PR, 10-14, para RT, e 18-21, para IN). Na generalidade, os produtos de PCR apresentavam-se específicos, com os tamanhos aproximados esperados para cada região em estudo. Assim, e sempre que se obtinham quantidades significativas de DNA, os produtos de PCR eram de imediato enviados para sequenciação nucleotídica directa (STAB VIDA, Caparica, Portugal), após purificação. Em algumas amostras, no entanto, foi observada a presença de fragmentos resultantes de amplificações inespecíficas (por exemplo, poço 4 da **Figura 3.3**). Nesses casos, e dependendo do padrão de inespecificidade, procedia-se a uma purificação dos fragmentos específicos recorrendo à extracção a partir do gel ou optava-se pela clonagem em vector plasmídico (ver secção 3.4).

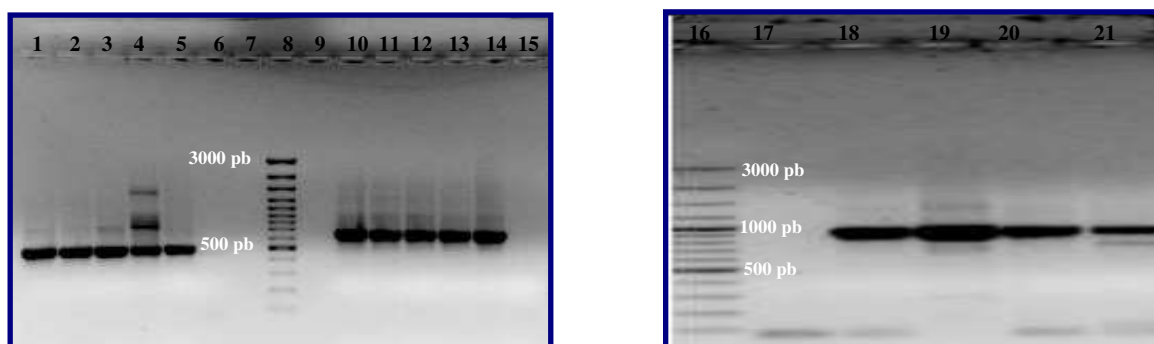


Figura 3.3. Visualização dos produtos de amplificação por RT-nested PCR das regiões da protease e transcriptase reversa (à esquerda) e da integrase (à direita) do gene *pol*, após electroforese em gel de agarose a 1,2% (p/v), em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml) (Anexo 2).

Identificação dos poços: 1 a 5: Amostras em estudo para a região da protease; 10 a 14: Amostras em estudo para a região da transcriptase reversa; 18 a 21: Amostras em estudo para a região da integrase; 6, 15 e 17: Controlo de contaminação (DNA matriz substituído por água); 8 e 16: Marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Canadá); 7 e 9: Pocos vazios.

3.3.2. Retrotranscrição *in vitro* com primers hexaméricos aleatórios

Todas as amostras que apresentaram um resultado negativo com o protocolo anterior (PT_TP04, PT_TP08, PT_TP42, PT_TP44, PT_TP47, PT_TP55, PT_TP63, PT_TP114, PT_TX20 e PT_TX43, para as três regiões em estudo, PT_TP85, somente para as regiões PR e IN, e PT_TX14, para a IN), foram novamente amplificadas por *nested* PCR, mas agora após um passo inicial de retrotranscrição *in vitro* com primers hexaméricos aleatórios.

Após observação do gel de agarose, verificou-se que foram amplificadas com êxito apenas três amostras, nomeadamente PT_TP114, para as três regiões em estudo (poços 9, 19 e 29 da **Figura 3.4**), PT_TP47, para a IN (poço 26 da **Figura 3.4**), e PT_TP85, para as regiões da PR e IN (**Figura 3.5**). Os fragmentos específicos de DNA foram enviados directamente para sequenciação nucleotídica (STAB VIDA, Caparica, Portugal), após purificação, utilizando o *QIAquick PCR Purification Kit* (QIAGEN, Alemanha).

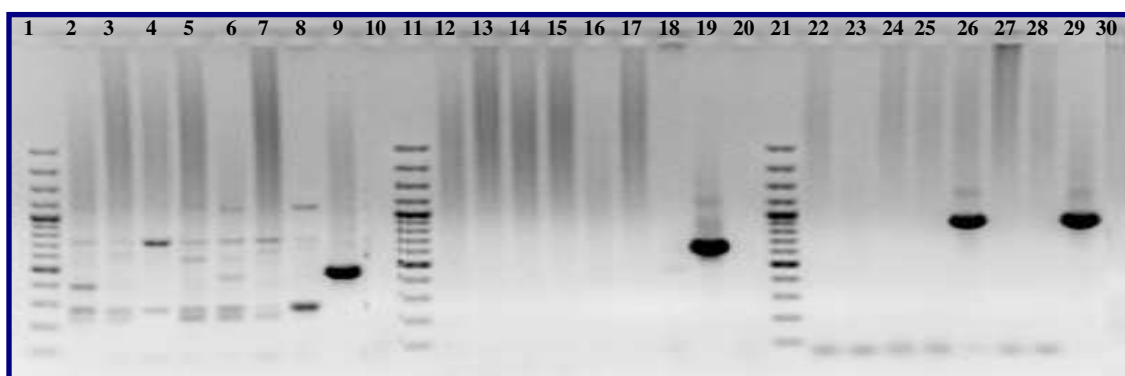


Figura 3.4. Visualização dos produtos de amplificação por RT-nested PCR das regiões da protease, transcriptase reversa e integrase do gene *pol*, após electroforese em gel de agarose a 1,2 % (p/v) em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml) (Anexo 2).

Identificação dos poços: 2 a 9: Amostras em estudo para a região da protease; 12 a 19: Amostras em estudo para a região da transcriptase reversa; 22 a 29: Amostras em estudo para a região da integrase; 10, 20 e 30: Controlo de contaminação (DNA matriz substituído por água); 1, 11 e 21: Marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Canadá).

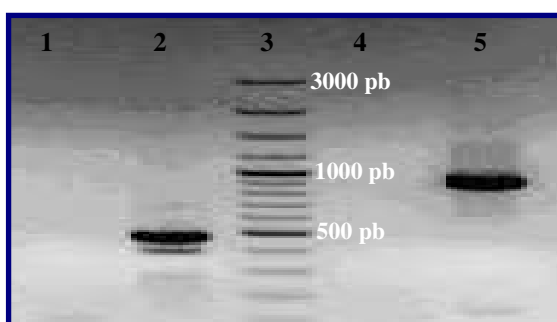


Figura 3.5. Visualização dos produtos de amplificação por RT-nested PCR das regiões da protease e integrase do gene *pol* (amostra PT_TP85), após electroforese em gel de agarose a 1,2% (p/v), em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml) (Anexo 2).

Identificação dos poços: 2 e 5: Regiões da protease e integrase, respectivamente; 1 e 4: Controlo de contaminação (DNA matriz substituído por água); 3: Marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Canadá).

É de salientar que o produto de amplificação da protease a partir da amostra PT_TP85 veio acompanhado de um fragmento de menores dimensões, resultante de amplificação inespecífica (**Figura 3.5**). Este produto foi enviado para sequenciação nucleotídica, após purificação do fragmento específico de DNA através da extracção a partir do gel, recorrendo ao sistema *innuPREP Gel Extraction Kit* (Analytik Jena AG, Alemanha).

Finalmente, como última tentativa, as amostras para as quais se tinha obtido produto de amplificação para apenas uma ou duas das três regiões em estudo, foram novamente submetidas a amplificação (após retrotranscrição *in vitro* com *primers* hexaméricos aleatórios). Para tal, procedeu-se à variação de algumas condições do protocolo de amplificação, inicialmente definidas:

o volume do produto de amplificação transferido da primeira para a segunda reacção de PCR (*nested* PCR) foi aumentado de 2 para 4 µl;

a hibridação dos *primers* utilizados na primeira reacção de amplificação (*nested* PCR), quer para a região da PR, quer para RT, foi realizada à temperatura única de 50°C (ver protocolo original na **Tabela 2.2**).

Após a implementação das alterações acima referidas, foi realizada uma electroforese em gel de agarose para a visualização dos produtos de amplificação, verificando-se a ausência de qualquer fragmento de DNA para as regiões em estudo.

No cômputo geral, e considerando os resultados obtidos pela utilização das duas metodologias de retrotranscrição *in vitro* (com *primers* específicos ou hexaméricos aleatórios), foi possível obter 42 amplicões para cada uma das três regiões (82,4% de positividade) (**Figura 3.6**), num total de 126 produtos de amplificação.

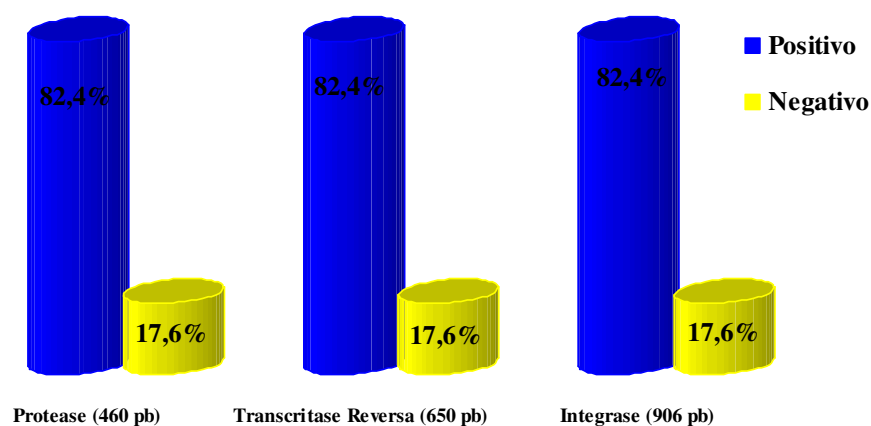


Figura 3.6. Representação gráfica dos resultados obtidos por RT-*nested* PCR (retrotranscrição *in vitro* com *primers* específicos ou hexaméricos aleatórios) para as três regiões do gene *pol* estudadas (n=51 amostras).

3.4. Clonagem no Vector Plasmídico *pGEM®-T Easy*

Do total de 126 produtos amplificados, correspondentes às regiões da PR, RT e IN, procedeu-se ao processo de clonagem para 15 deles (11,9%) [PR: PT_TP13,

PT_TX21, PT_TX50, PT_TX54, PT_TX59, PT_TX60; RT: PT_TP85, PT_TX48; IN: PT_TX13, PT_TX15, PT_TX16, PT_TX33, PT_TX34, PT_TX50, PT_TX59]. Para tal, recorreu-se ao sistema *pGEM[®]-T Easy Vector* (Promega, EUA), utilizando como hospedeiro *Escherichia coli* JM109. A clonagem foi realizada nas seguintes circunstâncias:

resultados de amplificação apresentando inespecificidade [com excepção das amostras PT_TP24 (poço 4 da **Figura 3.3**) e PT_TP85 (poço 2 da **Figura 3.5**), para a PR, para as quais foi realizada extracção e purificação dos produtos específicos directamente a partir do gel];

produtos amplificados com reduzida quantidade de DNA;

após sequenciação nucleotídica, cromatogramas difíceis de interpretar/editar, para os quais seria impossível obter uma sequência de qualidade.

Após transformação de células competentes de *E. coli* JM109 com as misturas de ligação, procedeu-se à selecção de clones bacterianos recombinantes (colónias de cor branca), por incubação em meio LB agar, na presença de ampicilina, X-gal e IPTG (**Anexo 2**), com base no sistema de α -complementação descrito no capítulo anterior (secção 2.9.1). As colónias de cor branca (**Figura 3.7**, como exemplo) eram repicadas, em condições de assepsia, para meio LB fresco com ampicilina e, após crescimento, lisadas para a extracção do DNA plasmídico pelo método de lise alcalina (secção 2.10).

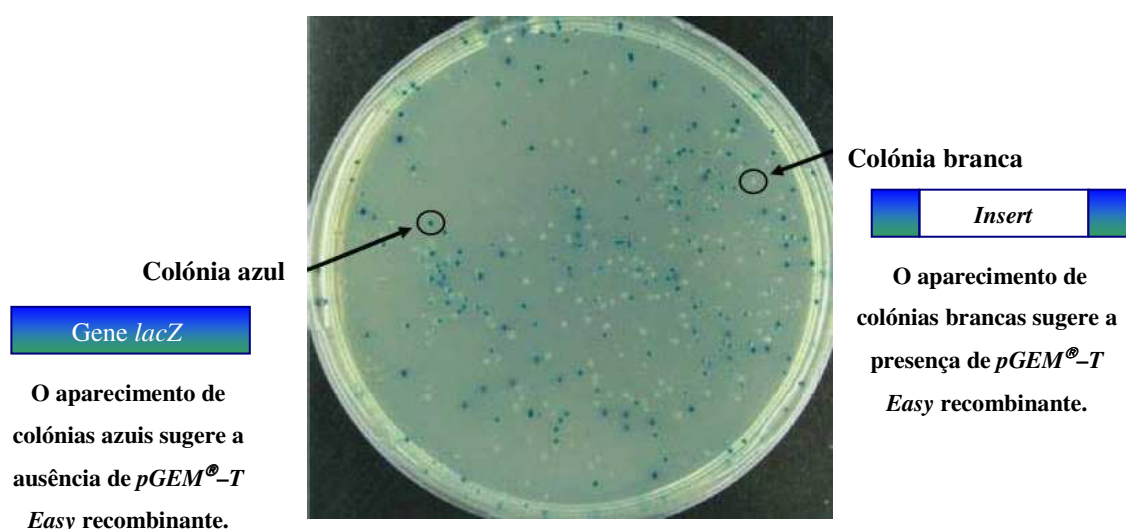


Figura 3.7. Rastreo de clones bacterianos recombinantes baseado no sistema de α -complementação branco/azul, após ligação dos produtos de PCR no vector plasmídico *pGEM[®]-T Easy*, transformação por choque térmico e plaqueamento em meio LB agar com ampicilina, X-gal e IPTG. O aparecimento de colónias brancas sugere a inserção do fragmento em estudo no local de clonagem múltipla no gene *lacZ*.

3.4.1. Selecção de plasmídeos recombinantes

Após a extracção do DNA plasmídico, as moléculas recombinantes foram identificadas pela análise do seu padrão de migração em campo eléctrico (exemplo, poços 1 a 16 ou A a U, **Figura 3.8**). A fim de se confirmar a correcta inserção do fragmento pretendido, procedeu-se à hidrólise dos DNAs plasmídicos apresentando migrações relativas mais retardadas, com a enzima de restrição EcoRI, uma vez que esta enzima reconhece locais, que flanqueiam o sítio de clonagem. Após hidrólise dos plasmídeos recombinantes putativos, foram observados em gel de agarose diversos padrões de restrição (exemplo, poços I a XI, **Figura 3.8**), entre os quais, alguns claramente indicadores da clonagem bem sucedida dos fragmentos de DNA pretendidos. O padrão de restrição mais comum era constituído por dois fragmentos de DNA, um de aproximadamente 3000 pb, correspondente ao vector plasmídico linearizado (3015 pb) e outro, consoante a região estudada, correspondente ao

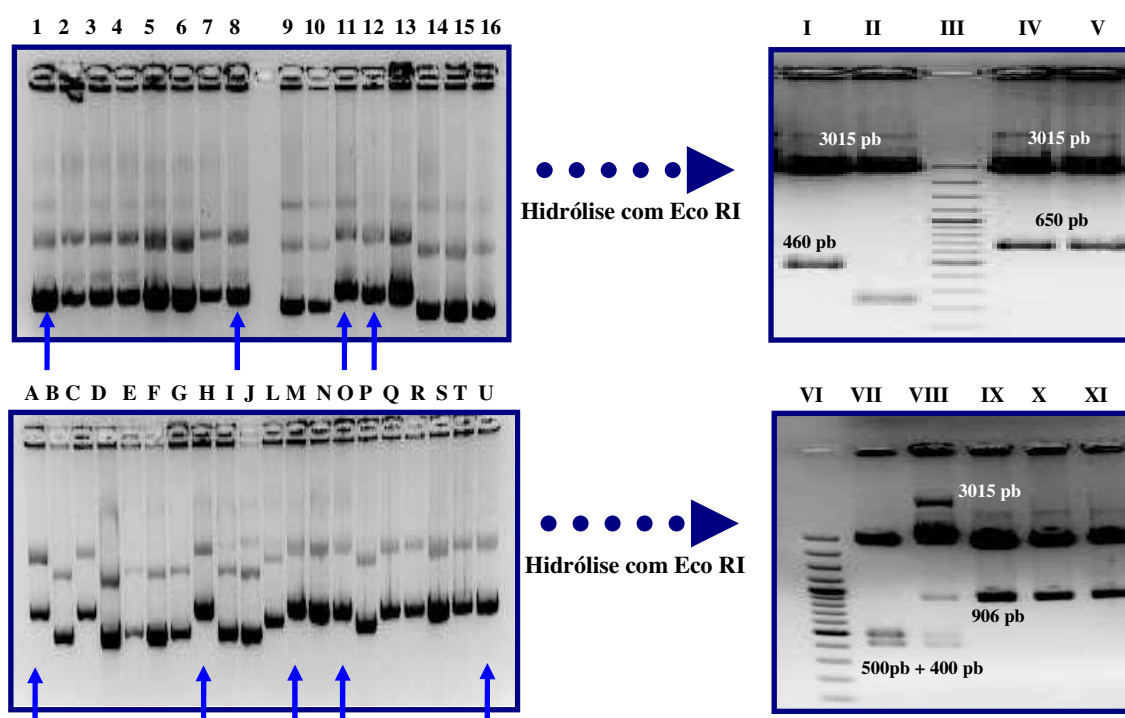


Figura 3.8. Visualização dos DNAs plasmídicos (à esquerda) e dos seus perfis de digestão enzimática com EcoRI (à direita), após electroforese em gel de agarose a 1,0-1,2% (p/v), em tampão TAE 0,5X, com brometo de etídeo (0,5 µg/ml) (Anexo 2). Os DNAs plasmídicos recombinantes putativos seleccionados para hidrólise enzimática encontram-se assinalados com setas de cor azul (à esquerda).

Identificação dos poços: 1 a 8, 9 a 16, A a M e N a U: DNAs plasmídicos de quatro amostras diferentes; I e II: Perfis de restrição para PR (clone positivo e negativo, respectivamente); IV e V: Perfis de restrição para RT (clones positivos); VII a XI: Perfis de restrição distintos para IN (clones positivos); III e VI: Marcador de massa molecular *GeneRuler™ 100 bp Plus DNA Ladder* (Fermentas, Canadá).

fragmento de DNA clonado de 460 pb (poço I da **Figura 3.8**), 650 pb (poços IV e V da **Figura 3.8**) ou 906 pb (poços IX a XI da **Figura 3.8**). É de realçar que para as regiões da RT e IN foi identificado, em alguns clones, um segundo local de reconhecimento para a enzima de restrição EcoRI. Nesses casos, três fragmentos de DNA foram observados após restrição, um correspondente ao vector plasmídico linearizado (~3000 pb) e dois fragmentos adicionais, com tamanho aproximado de 230 e 420 pb (resultados não apresentados), relativos ao fragmento clonado da RT, ou de aproximadamente 400 e 500 pb, relativos ao fragmento clonado da IN (exemplo, poços VII e VIII, este evidenciando digestão parcial, na **Figura 3.8**).

3.5. Análise Bioinformática de Sequências Nucleotídicas

3.5.1. Análise dos cromatogramas e edição das sequências

Após obtenção dos cromatogramas, relativos às reacções de sequenciação das cadeias directa e complementar de DNA com os respectivos *primers* (referidos no capítulo 2.5), referentes aos 126 fragmentos amplificados das três regiões em estudo do gene *pol* de HIV-1, procedeu-se à edição das respectivas sequências nucleotídicas. Para tal, os cromatogramas de cada amplicão foram cuidadosamente comparados e analisados a fim de se obter uma sequência “consenso” entre ambos. Eventuais discordâncias entre as sequências complementares eram corrigidas manualmente, após confirmação das posições em causa nos respectivos cromatogramas. Nas discrepâncias impossíveis de resolver, optou-se pela utilização do código degenerado de nucleótidos (disponível em www.promega.com.techserv). Em suma, dos 126 fragmentos de DNA enviados para sequenciação foram obtidas 123 sequências nucleotídicas de “consenso” (97,6%) (**Tabela 3.2**). A edição de 2,4% das sequências (3/126) revelou-se impossível, por uma das seguintes razões:

- qualidade inaceitável do cromatograma, caracterizada pela presença de “ruído de fundo” excessivo;
- cromatogramas de difícil interpretação, devido à presença de elevado número de degenerações;
- obtenção de sequências nucleotídicas truncadas.

Tabela 3.2. Número total de sequências nucleotídicas editadas e não editadas no presente estudo.

Número total de fragmentos de DNA sequenciados	Número total de sequências editadas			Número total de sequências não editadas		
126	123 (97,6%)			3 (2,4%)		
	PR	RT	IN	PR	RT	IN
	41	41	41	1	1	1

Após alinhamento múltiplo das sequências pré-editadas, contra a sequência de referência HXB2 de HIV-1, procedeu-se à eliminação das porções de nucleótidos que não correspondiam às regiões codificantes PR, RT ou IN, propriamente ditas (embora pudessem ter sido amplificadas) (Tabela 3.3).

Tabela 3.3. Tamanho do fragmento amplificado por *nested* PCR, tamanho das sequências editadas e respectiva localização em HXB2 e tamanho da região estudada relativamente a HXB2.

Regiões estudadas	Tamanho do fragmento amplificado por <i>nested</i> PCR (pb)	Tamanho das sequências editadas (pb) e localização em HXB2	Proporção estudada relativamente a HXB2 (%)
Protease	460	297 [2253-2549]	100 [297/297]
Transcritase Reversa	650	598 [2633-3230]	35,6 [598/1680]
Integrase	906	813 [4230-5042]	94,1 [813/864]

No cômputo geral, considerando as três regiões estudadas, no seu conjunto, foi possível obter, neste trabalho, cerca de 60,1% (1708/2841) da região codificante do gene *pol* de HIV-1.

3.5.2. Pesquisa de homologia genética

Para cada uma das 123 sequências editadas, foi realizada uma pesquisa automática e rápida de homologia genética, recorrendo-se ao programa *BLAST* (disponível em <http://blast.ncbi.nlm.nih.gov/Blast.cgi>), a fim de se confirmar a sua identidade. Desta pesquisa foi obtido um relatório final que lista as sequências nucleotídicas da base de dados *GenBank/EMBL/DDBJ* que possuem a maior similaridade com cada uma das sequências em estudo e a cada uma das comparações estabelecidas encontra-se associada uma pontuação (*Total score*) que quantifica o grau

de semelhança entre as duas sequências nucleotídicas. Para dar credibilidade aos resultados obtidos, o programa calcula um parâmetro estatístico denominado *E value*. Este valor, variando entre zero e um, corresponde à probabilidade de uma sequência aleatória, com o mesmo tamanho e composição, originar um alinhamento com um *score* igual ou superior, ou, por outras palavras, estima a probabilidade da semelhança genética encontrada ser devida ao acaso. Desta forma, quanto mais próximo de zero for o *E value*, mais fiável é a homologia encontrada. Para todas as sequências em estudo, foram observados *scores* elevados, associados a valores de *E value* igual a zero, e percentagens de identidade igual ou superior a 96% com sequências nucleotídicas de HIV-1 presentes na base de dados. Adicionalmente, foi realizada uma pesquisa automática de homologia genética recorrendo ao programa *NCBI HIV Genotyping Program* (disponível em <http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>). Esta ferramenta, baseada no *BLAST*, permite a pesquisa de regiões de similaridade local entre a sequência em estudo e sequências de HIV de referência, existentes numa base de dados específica, permitindo, de forma rápida, contudo pouco robusta, o reconhecimento de eventuais formas virais recombinantes.

Em resumo, com esta análise inicial foi possível confirmar a obtenção de sequências da PR, RT e IN do gene *pol* do HIV-1, simultaneamente com o reconhecimento do grau de semelhança genético relativo aos diferentes subtipos/CRFs virais (resultados não apresentados).

3.5.3 Análise filogenética

3.5.3.1 Determinação automática de genótipos virais

Procedeu-se à subtipagem preliminar das sequências em estudo, sendo esta realizada por recurso ao programa *REGA HIV-1 & 2 Automated Subtyping Tool*, v.2.0 (disponível em <http://jose.med.kuleuven.be/genotypetool/html/subtypinghiv.html>). Este programa de subtipagem baseia-se num algoritmo implementado de modo automático que permite a realização de uma análise filogenética rápida suplementada com métodos de *bootstrapping* e também de *bootscanning* para pesquisa de formas recombinantes (de salientar que, neste programa, a análise de *bootscanning* é possível somente para sequências de tamanho igual ou superior a 800 pb, o que excluiu, desde logo, desta análise, as sequências PR e RT). As análises (filogenética e de *bootscanning*) realizadas

para as sequências em estudo utilizaram um conjunto de sequências de referência de subtipos puros e de formas recombinantes circulantes de HIV-1 do grupo M. O relatório final apresentado pelo programa, para além dos resultados da análise filogenética propriamente dita, indica também o “sinal filogenético”, o qual permite inferir sobre a qualidade do alinhamento múltiplo realizado para a construção da árvore filogenética apresentada. Os resultados obtidos nesta análise filogenética automática são apresentados na **Tabela 3.4.** (apenas valores de *bootstrap* superiores a 70% são considerados significativos pelo programa).

Tabela 3.4. Resultados da determinação automática dos genótipos virais [*REGA HIV-1 & 2 Automated Subtyping Tool*, v.2.0]

Genótipo	PR (n=41)	RT (n=41)	IN (n=41)
B	13 (31,7%)	16 (39,0%)	13 (31,7%)
F (F1)	0	1 (2,4%)	1 (2,4%)
G	16 (39,0%)	20 (48,8%)	13 (31,7%)
G(14_BG)	---*	---*	5 (12,2%)
CRF02_AG	---*	---*	3 (7,3%)
Recombinantes intra-região	---*	---*	6 (14,6%)
Não atribuído	12 (29,3%)	4 (9,8%)	0

*Neste programa, a análise de *bootscanning* é possível somente para sequências de tamanho igual ou superior a 800 pb, o que excluiu, desde logo, as sequências PR e RT

Como se pode observar na **Tabela 3.4**, o programa *REGA HIV-1 & 2 Automated Subtyping Tool* v.2.0 não atribuiu um genótipo a 12 sequências PR (PT_TP19, PT_TP71, PT_TP107, PT_TP114, PT_TX13, PT_TX16, PT_TX33, PT_TX41, PT_TX46, PT_TX47, PT_TX52 e PT_AS646) e a 4 sequências RT (PT_TP24, PT_TX14, PT_TX18 e PT_TX40). Pela observação das árvores filogenéticas construídas para cada uma destas sequências constatou-se uma das seguintes situações:

forma um grupo monofilético com as sequências de referência de um subtipo puro do HIV-1 do grupo M, mas suportado por um valor de *bootstrap* não significativo (<70%);

forma um grupo monofilético com as sequências de referência de um subtipo puro do HIV-1 do grupo M, suportado por um valor de *bootstrap* significativo (>70%),

mas o ramo que a suporta evidencia um posicionamento divergente precoce, relativamente ao agrupamento definido pelas referências.

3.5.3.2 Determinação de genótipos virais após construção de árvores filogenéticas

Os resultados apresentados na **Tabela 3.4** foram confirmados e/ou complementados através da construção de árvores filogenéticas, recorrendo a métodos não automatizados que permitem a selecção, entre outros, dos algoritmos e das sequências de referência a utilizar, constituindo, por isso, métodos filogeneticamente mais robustos. Para além da construção de árvores filogenéticas para cada uma das regiões codificantes do gene *pol* do HIV-1 (PR, RT e IN), procedeu-se ainda à construção de uma árvore baseada numa sequência artificial concatenada, resultante da simples justaposição das três sequências individuais, a fim de aumentar a informação em análise. É de realçar que após a realização dos alinhamentos múltiplos necessários para a construção das árvores filogenéticas, para o qual se recorreu à ferramenta bioinformática *ClustalW2* (218, disponível em www.ebi.ac.uk/clustalw/), e durante a edição manual no programa *BioEdit v.7.0.5.3.* (87), não foram observadas deleções ou inserções que originassem perda de informação nas três regiões em estudo. Somente no alinhamento relativo à sequência concatenada foi observada a presença de uma deleção de um nucleótido (numa sequência de referência do subtipo J, posição 435), o que levou à eliminação da coluna correspondente à respectiva posição, passando o alinhamento a conter um total de 1707 nucleótidos.

Tal como descrito anteriormente, a construção de todas as árvores filogenéticas foi efectuada recorrendo ao programa *MEGA v.4.0* (214, disponível em <http://www.megasoftware.net/>), com recurso ao método de *Neighbor-Joining* (184), baseado no modelo de dois parâmetros de Kimura para o cálculo das matrizes de distância (108). A robustez da inferência filogenética foi testada por análise de *bootstrapping* de 1000 replicados aleatórios, considerando significativos valores >70%. O número total de sequências (em estudo e de referência) utilizadas na construção de cada árvore encontra-se na **Tabela 3.5**.

Tabela 3.5. Número de sequências nucleotídicas utilizadas na construção das árvores filogenéticas.

Árvore filogenética	Sequências em estudo	Sequências de referência	Total
Protease	41	53	94
Transcritase Reversa	41	52	93
Integrase	41	39	80
Concatenada	38*	38	76

*Número de amostras para as quais foram obtidas sequências nucleotídicas das três regiões codificantes do gene *pol* do HIV-1

Após construção das árvores filogenéticas correspondentes às três regiões codificantes do gene *pol* foi verificado que as sequências nucleotídicas em estudo formavam grupos monofiléticos, em geral, bem definidos, com as respectivas referências do HIV-1. Por outro lado, foi observado que a grande maioria dos agrupamentos formados para as sequências correspondentes às regiões RT e IN eram suportadas por valores de *bootstrap* muito significativos (superiores a 70%). Assim, para as regiões PR (**Figura 3.9A**), RT (**Figura 3.10A**) e IN (**Figura 3.11A**) foi possível identificar a maioria dos genótipos das sequências em estudo (**Anexo 4**). Contudo, é de salientar que a árvore filogenética construída para a região da PR apresenta agrupamentos suportados com valores de *bootstrap*, em regra, inferiores a 70%, impedindo que a classificação do genótipo das sequências em estudo fosse realizada com um elevado grau de certeza. Exceptua-se a sequência PT_AS646, agrupada com sequências de referência do sub-subtipo F1, com um valor de *bootstrap* de 95% (**Figura 3.9A**). Em termos globais, a análise filogenética realizada revelou os resultados que se encontram resumidos na **Tabela 3.6**.

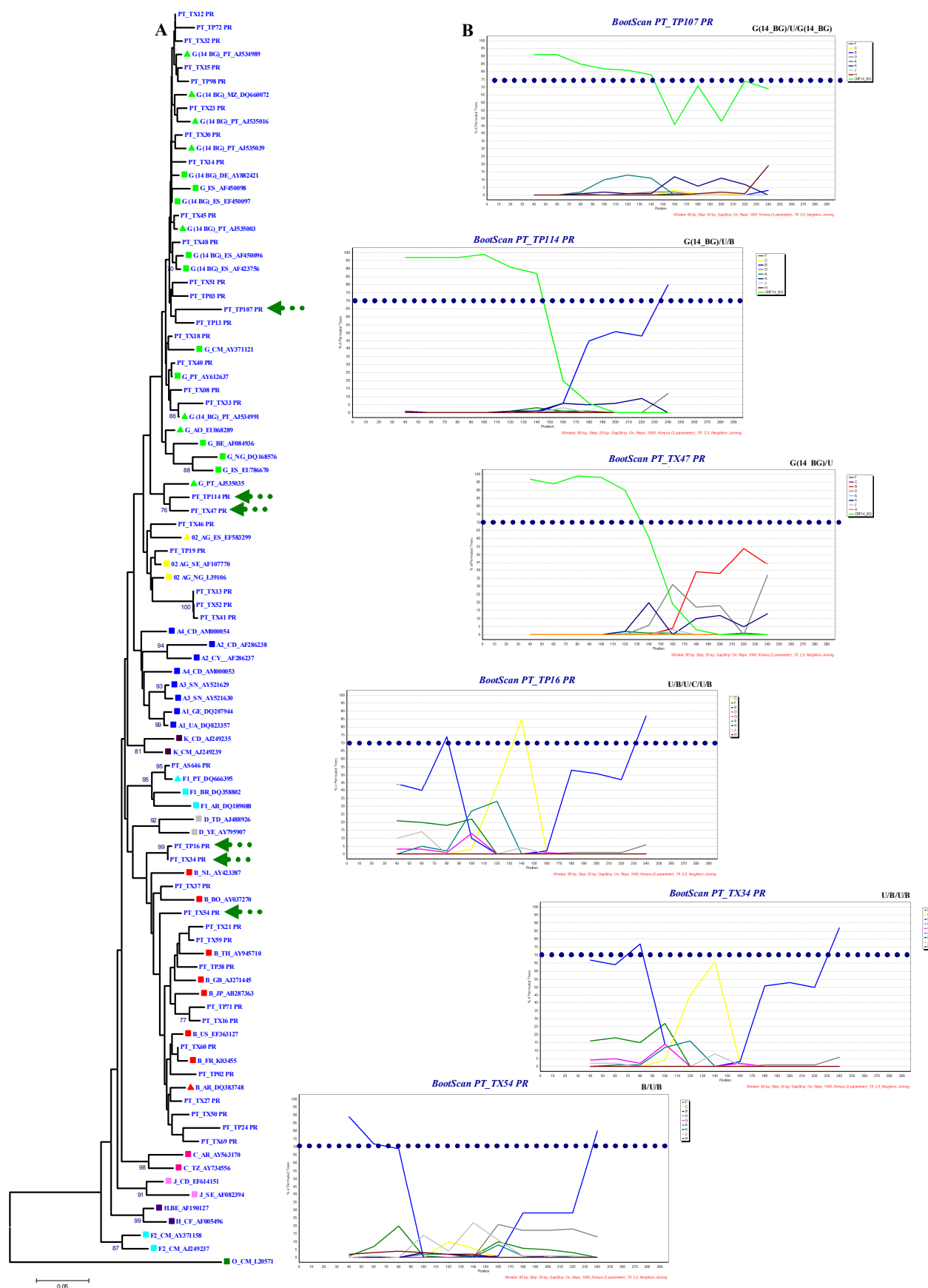
Análise Filogenética e de Recombinantes da Região da Protease do Gene *pol* do HIV-1

Figura 3.9. A) Relações filogenéticas entre 94 sequências nucleotídicas correspondentes à região da protease do gene *pol* do HIV-1, a partir de um alinhamento de 297 nucleótidos. Na árvore filogenética estão representadas 41 sequências em estudo (assinaladas com o prefixo PT), 42 sequências de referência dos grupos M (AI-A4, B, C, D, F1, F2, G, H, J, K, CRF02_AG e CRF14_BG) e O (utilizada como *outgroup*), extraídas da base de dados *Los Alamos HIV Database* (representadas com um quadrado colorido) e 11 sequências nucleotídicas obtidas após pesquisa de homologia por *BLAST* contra as sequências em estudo (representadas com um triângulo colorido). Cada sequência de referência é precedida de uma letra correspondente ao respectivo genótipo, seguida de uma abreviatura para o local de origem (AO- Angola, AR- Argentina, BE- Bélgica, BO- Bolívia, BR- Brasil, CD- República Democrática do Congo, CF- República Centro-Africana, CM- Camarões, CY- Chipre, DE- Alemanha, ES- Espanha, FR- França, GB- Grã-Bretanha, GE- Geórgia, JP- Japão, MZ- Moçambique, NG- Nigéria, NL- Holanda, PT- Portugal, SE- Suécia, SN- Senegal, TD- Chade, TH- Tailândia, TZ- Tanzânia, UA- Ucrânia, US- Estados Unidos da América, YE- Iêmen) e, finalmente, do número de acesso do isolado viral. A análise filogenética foi conduzida no programa *MEGA 4 v. 4.0* e a história evolutiva baseada no método filogenético *Neighbor-Joining* baseado no modelo 2 parâmetros de Kimura para o cálculo da matriz de distâncias genéticas (a escala indica 5% de diversidade genética). A robustez da inferência filogenética foi avaliada por *bootstrapping* de 1000 replicados, considerando-se representativos para genótipos valores iguais ou superiores a 70%. B) À direita da árvore filogenética encontram-se representados os resultados da análise de *bootcanning* (janela 80 bp; passo 20 bp) por *SimPlot v. 3.5.1* das 6 sequências assinaladas (com uma seta verde). As linhas horizontais pontilhadas indicam o *cut-off* de 70%.

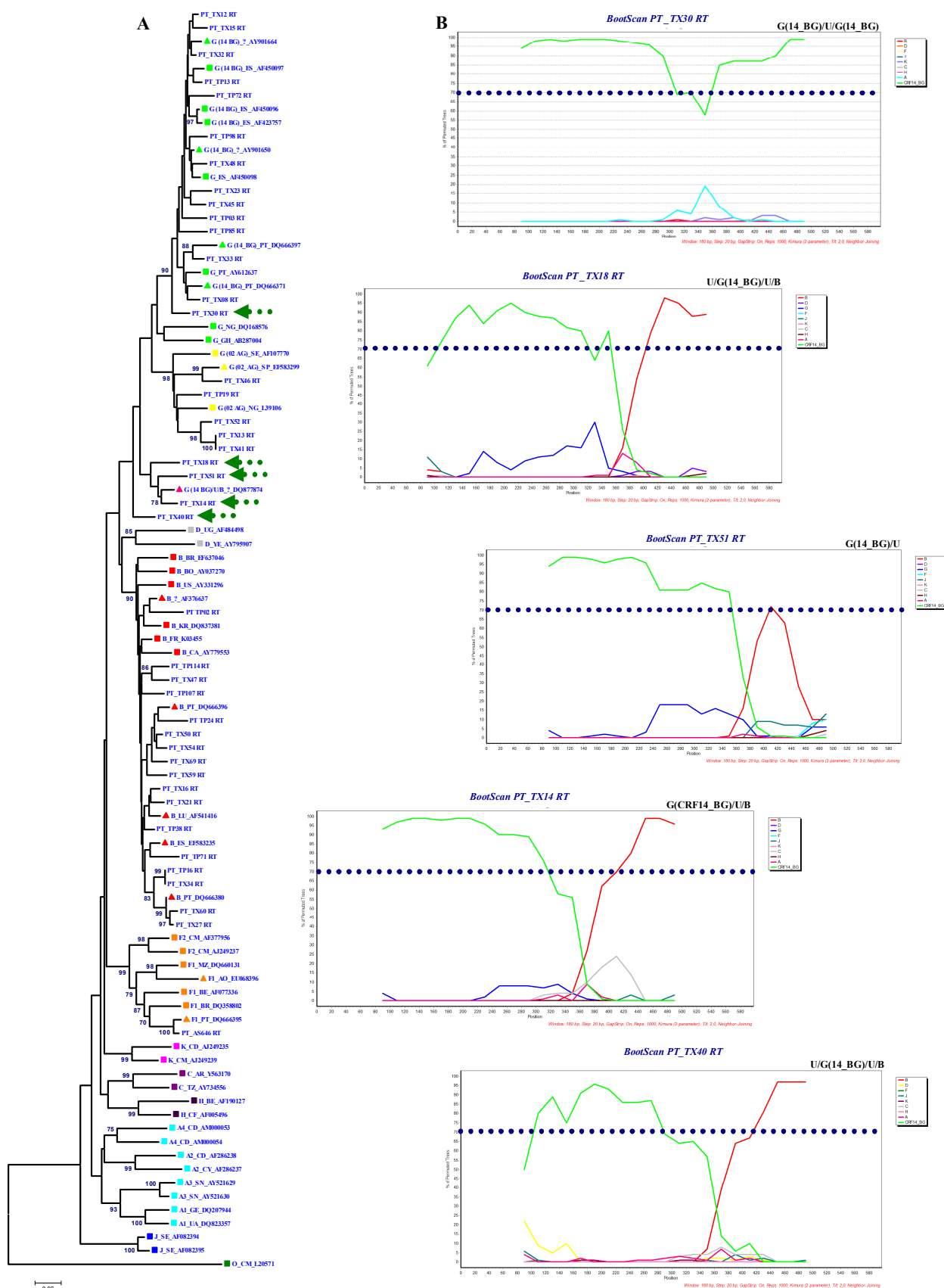
Análise Filogenética e de Recombinantes da Região da Transcriptase Reversa do Gene *pol* do HIV-1

Figura 3.10. A) Relações filogenéticas entre 93 sequências nucleotídicas correspondentes à região da transcriptase reversa do gene *pol* do HIV-1, a partir de um alinhamento de 598 nucleótidos. Na árvore filogenética estão representadas 41 sequências em estudo (assinadas com o prefixo PT), 39 sequências de referência dos grupos M (A1-A4, B, C, D, F1, F2, G, H, J, K, CRF02_AG e CRF14_BG) e O (utilizada como *outgroup*), extraídas da base de dados *Los Alamos HIV Database* (representadas com um quadrado colorido) e 13 sequências nucleotídicas obtidas após pesquisa de homologia por *BLAST*, contra as sequências em estudo (representadas com um triângulo colorido). Cada sequência de referência é precedida de uma letra correspondente ao respectivo genótipo, seguida de uma abreviatura para o local de origem (AO- Angola, AR- Argentina, BE- Bélgica, BO- Bolívia, BR- Brasil, CA- Canadá, CD- República Democrática do Congo, CF- República Centro - Africana, CM- Camarões, CY- Chipre, ES- Espanha, FR- França, GE- Geórgia, GH- Gana, KR- Coreia do Sul, LU- Luxemburgo, MZ- Moçambique, NG- Nigéria, PT- Portugal, SE- Suécia, SN- Senegal, TZ- Tanzânia, UA- Ucrânia, UG- Uganda, US- Estados Unidos da América, YE- Iémen e ?- Origem desconhecida) e, finalmente, do número de acesso do isolado viral. A análise filogenética foi conduzida no programa *MEGA4 v. 4.0* e a história evolutiva baseada no método filogenético *Neighbor-Joining* e no modelo 2 parâmetros de Kimura para o cálculo da matriz de distâncias genéticas (a escala indica 2% de diversidade genética). A robustez da inferência filogenética foi avaliada por *bootstrapping* de 1000 replicados, considerando-se representativos para genótipos valores iguais ou superiores a 70%. B) À direita da árvore filogenética encontram-se representados a análise de *bootscanning* (janela 180 bp; passo 20 bp) por *SimPlot v. 3.5.1* das 5 sequências assinadas (com uma seta verde). As linhas horizontais pontilhadas indicam o *cut-off* de 70%.

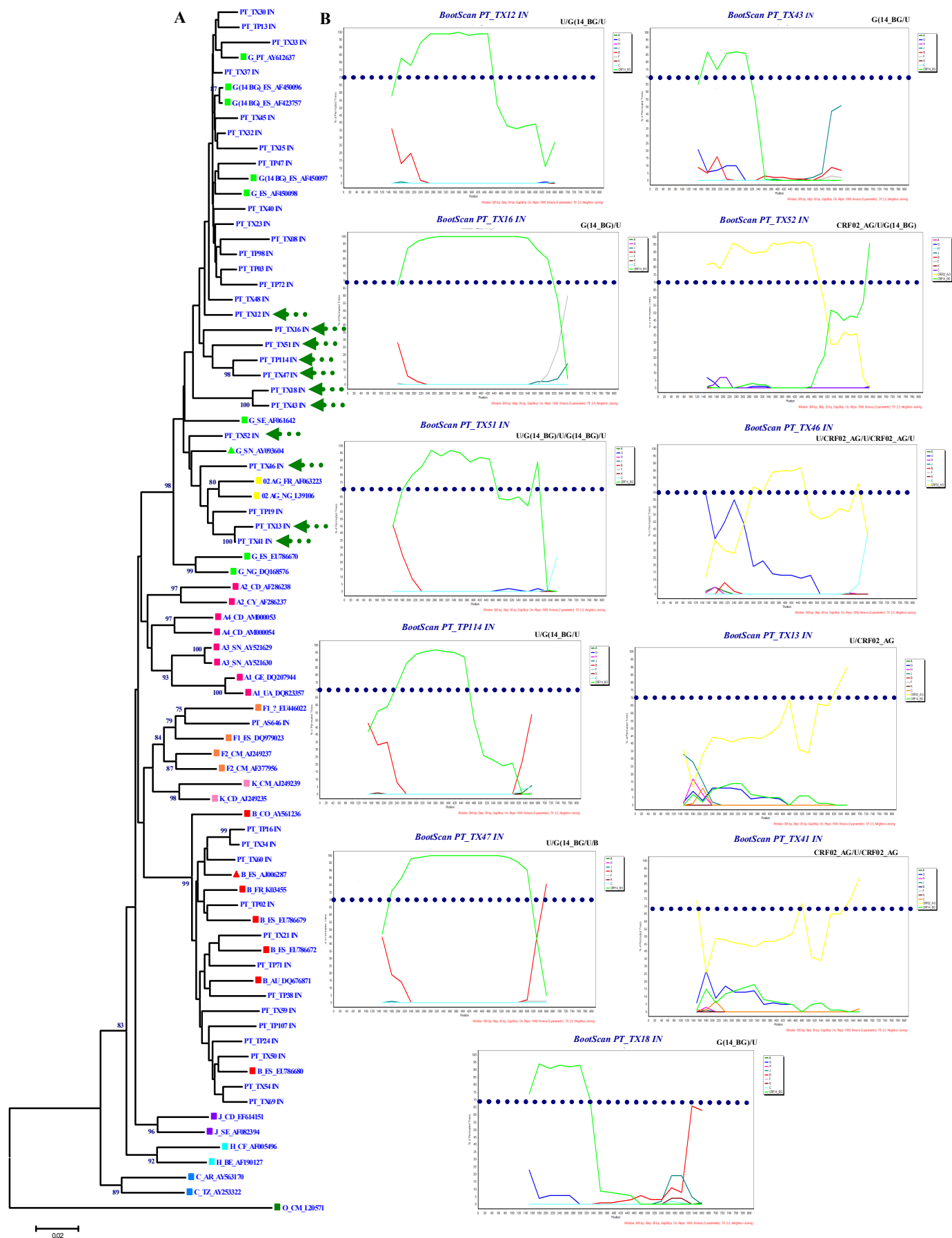
Análise Filogenética e de Recombinantes da Região da Integrase do Gene *pol* do HIV-1

Figura 3.11. A) Relações filogenéticas entre 80 sequências nucleotídicas correspondentes à região da integrase do gene *pol* do HIV-1, a partir de um alinhamento de 813 nucleótidos. Na árvore filogenética estão representadas 41 sequências em estudo (assinadas com o prefixo PT), 37 sequências de referência dos grupos M (A1-A4, B, C, D, F1, F2, G, H, J, K, CRF02_AG e CRF14_BG) e O (utilizada como *outgroup*), extraídas da base de dados *Los Alamos HIV Database* (representadas com um quadrado colorido) e 2 sequências nucleotídicas obtidas após pesquisa de homologia por *BLAST*, contra as sequências em estudo (representadas com um triângulo colorido). Cada sequência referência é precedida de uma letra correspondente ao respectivo genótipo, seguida de uma abreviatura para o local de origem (AR- Argentina, AU- Austrália, BE- Bélgica, CD- República Democrática do Congo, CF- República Centro - Africana, CM- Camarões, CO- Colômbia, CY- Chipre, ES- Espanha, FR- França, GE- Geórgia, NG- Nigéria, PT- Portugal, SE- Suécia, SN- Senegal, TZ- Tanzânia, UA- Ucrânia e ?- Origem desconhecida) e, finalmente, do número de acesso do isolado viral. A análise filogenética foi conduzida no programa *MEGA v. 4.0* e a história evolutiva baseada no método filogenético *Neighbor-Joining* e no modelo 2 parâmetros de Kimura para o cálculo da matriz de distâncias genéticas (a escala indica 2% de diversidade genética). A robustez da inferência filogenética foi avaliada por *bootstrapping* de 1000 replicados, considerando-se representativas para genotipagem valores iguais ou superiores a 70%. B) À direita da árvore filogenética encontram-se representados os resultados da análise de *bootscanning* (janela 300 bp; passo 30 bp) por *SimPlot v. 3.5.1* das 11 sequências assinadas (com uma seta verde). As linhas horizontais pontilhadas indicam o *cut-off* de 70%.

Tabela 3.6. Resultados de genotipagem após construção de árvores filogenéticas.

Genótipo provável	PR (n=41)	RT (n=41)	IN (n=41)
B	12 (29,3%)	17 (41,5%)	13 (31,7%)
F (F1)	1 (2,4%)	1 (2,4%)	1 (2,4%)
G ou G(14_BG)	17 (41,5%)	13 (31,7%)	12 (29,3%)
CRF02_AG ou G(02_AG)	5 (12,2%)	5 (12,2%)	1 (2,4%)
Não atribuído*	6 (14,6%)	5 (12,2%)	14 (34,2%)

*Corresponde a sequências com posicionamento ambíguo na árvore filogenética ou demonstrando elevada divergência genética, seleccionadas para análise de recombinantes por *bootscanning*

Comparando os resultados das análises filogenéticas das três regiões em estudo realizadas no programa *MEGA v.4.0* (**Tabela 3.6**) com os obtidos previamente com recurso ao programa *REGA HIV-1 & 2 Automated Subtyping Tool v.2.0* (**Tabela 3.4**), verificou-se um grau de concordância muito elevado (dados não apresentados). Contudo, entre os resultados obtidos pelos dois programas foram encontradas algumas diferenças, as quais serão abordadas em seguida. Relativamente a nove sequências da PR (PT_TP71, PT_TX16, PT_TX33, PT_TP19, PT_TX13, PT_TX52, PT_TX41, PT_TX46 e PT_AS646), não subtipadas recorrendo à genotipagem automática (**Tabela 3.4**) foi verificado, após construção da árvore filogenética (**Figura 3.9A**), que as duas primeiras partilhavam o mesmo ancestral que as referências do subtipo B e a terceira com as referências do subtipo G/G(14_BG), as cinco seguintes com a forma recombinante CRF02_AG e a última formava um agrupamento monofilético robusto com sequências de referência do sub-subtipo F1, suportado por um valor de *bootstrap* de 95%. Quanto às sequências da RT, as amostras PT_TX51 e PT_TX30 foram previamente genotipadas como subtipo G, mas na árvore filogenética apresentam ramos de suporte em posição ambígua, precocemente divergente e/ou com elevada divergência genética, sugerindo que possam ser, na realidade, recombinantes ou variantes genéticas do subtipo G (**Figura 3.10A**). Finalmente, no que se refere às sequências da IN (**Figura 3.11A**), o mesmo acontece para as amostras PT_TX08, PT_TX33, PT_TX48 e PT_TX52, classificadas anteriormente como G, e PT_TX12, genotipada como G(14_BG). Por último, no que respeita às sequências PT_TX13, PT_TX41 e PT_TX46 [previamente classificadas como CRF02_AG (as duas primeiras) e G (a última)], o seu

posicionamento na árvore filogenética sugere que possam tratar-se de formas CRF02_AG geneticamente divergentes.

Durante a análise das árvores filogenéticas foi constatado que algumas sequências (**Figura 3.12**) apresentavam posicionamentos ambíguos, dificultando a sua classificação (**Figuras 3.9A-3.11A**), podendo isto ser devido, como se viu, a fenômenos de recombinação (inter e intragenotípica) ou de hipermutação por ocorrência de hiperadenilação G→A (181). A confirmação de possíveis eventos de recombinação foi realizada por análise de *bootscanning* recorrendo-se ao programa *SimPlot v.3.5.1*. (120).

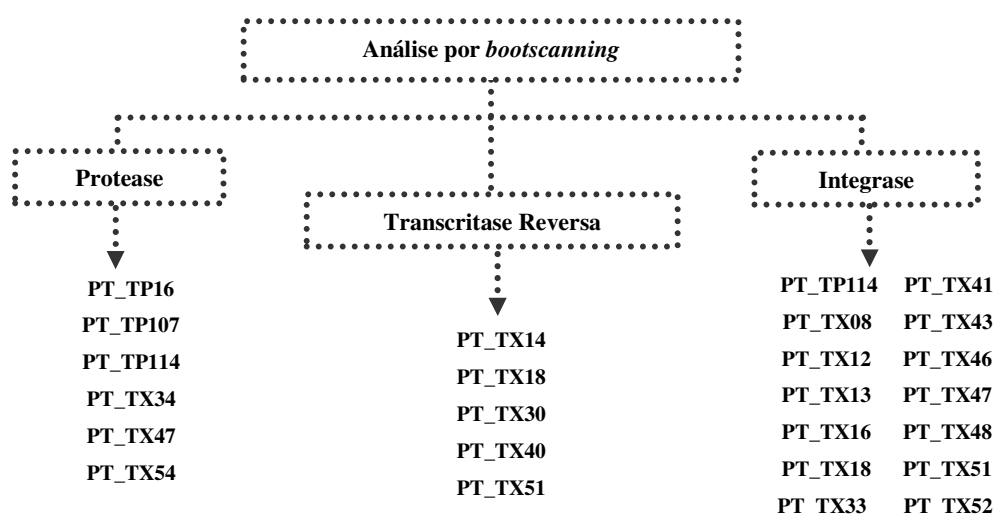


Figura 3.12. Identificação das sequências que apresentavam posicionamentos ambíguos na respectiva árvore filogenética dificultando a sua genotipagem.

A análise de *bootscanning* foi inicialmente realizada para cada uma das três regiões, tendo-se obtido os seguintes resultados:

Região da Protease

As sequências PT_TP16, PT_TX34 e PT_TX54 revelaram homologia parcial com o subtipo B, apresentando também segmentos não genotipáveis, ditos U (do inglês *unknown*). Por sua vez, as sequências PT_TP107, PT_TP114 e PT_TX47 apresentaram uma homologia parcial com o G(14_BG), acompanhada de regiões U (**Figura 3.9A**). De realçar que a significância deste tipo de análise para a PR é limitada, considerando que se está a analisar um fragmento de pequeno tamanho (297 nucleótidos).

Região da Transcriptase Reversa

As sequências PT_TX14, PT_TX18 e PT_TX40 foram identificadas como recombinantes intra-RT, sendo caracterizadas pela estrutura mosaico genérica G(14_BG)/B. A sequência PT_TX51 tratar-se-á também de um recombinante intra-RT, o qual demonstrou homologia parcial com o G(14_BG), tendo no final uma região U. Finalmente, a sequência PT_TX30 não revelou indícios de recombinação, possuindo homologia, em quase toda a sua extensão, com o G(14_BG), excepto num pequeno trecho central, no qual o sinal filogenético deixa de ser significativo (<70%). Esse facto explicará a sua posição na árvore filogenética, na qual o ramo que a suporta evidencia um posicionamento divergente precoce, relativamente ao agrupamento definido pelas referências, indicando a presença de uma estirpe viral divergente da forma recombinante circulante 14_BG (**Figura 3.10A**).

Região da Integrase

Por último, das 14 sequências correspondentes à região da IN submetidas a análise de *bootscanning*, PT_TX08, PT_TX33 e PT_TX48 apresentaram, em toda a sua extensão, homologia significativa com G(14_BG). Considerando também o seu posicionamento na árvore filogenética, pode colocar-se a hipótese de serem estirpes virais divergentes da forma recombinante circulante 14_BG. No que respeita ao grupo de sequências constituído por PT_TP114, PT_TX12, PT_TX16, PT_TX18, PT_TX43, PT_TX47 e PT_TX51, a análise de *bootscanning* revelou uma estrutura genérica G(14_BG)/U. Por outro lado, PT_TX13, PT_TX41 e PT_TX46 apresentaram homologia parcial, embora com uma significância relativamente reduzida, com CRF02_AG, revelando também regiões do genoma não genotipáveis. Finalmente, a sequência PT_TX52 apresentou uma estrutura recombinante CRF02_AG/G(14_BG) (**Figura 3.11A**).

A genotipagem realizada a partir da árvore filogenética construída com as sequências concatenadas (1707 nucleótidos), como se esperava, foi genericamente concordante com a análise anterior para as três regiões codificantes, individualmente, ajudando, no entanto, a definir melhor eventuais padrões de recombinação ou a existência de sequências geneticamente divergentes (**Anexo 4**).

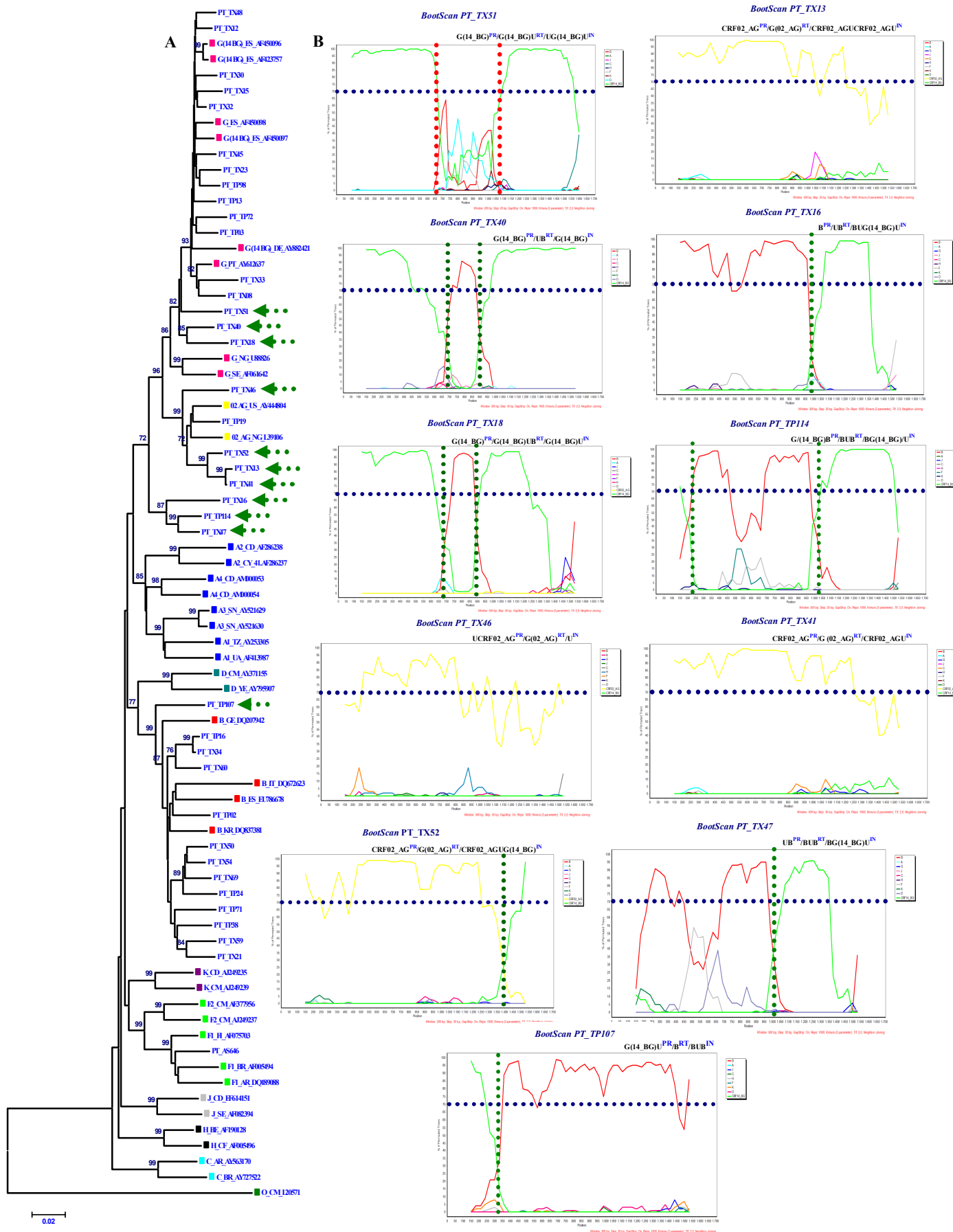
Análise Filogenética e de Recombinantes das Sequências Concatenadas (Protease+ Transcriptase Reversa+ Integrase do Gene *pol* do HIV-1

Figura 3.13. A) Relações filogenéticas entre 76 sequências nucleotídicas correspondentes às sequências concatenadas (PR+RT-IN), do gene *pol* do HIV-1, a partir de um alinhamento de 1707 nucleótidos. Na árvore filogenética estão representadas 38 sequências em estudo (assinadas com o prefixo PT), 38 sequências de referência dos grupos M (A1-A4, B, C, D, F1, F2, G, H, J, K, CRF02_AG e CRF14_BG) e O (utilizada como *outgroup*), extraídas da base de dados *Los Alamos HIV Database* (representadas com um quadrado colorido). Cada sequência de referência é precedida de uma letra correspondente ao respectivo genótipo, seguida de uma abreviatura para o local de origem (AR- Argentina, BE- Bélgica, BR- Brasil, CD- República Democrática do Congo, CF- República Centro - Africana, CM- Camarões, CY- Chipre, DE- Alemanha, ES- Espanha, FI- Finlândia, GE- Geórgia, IT- Itália, KR- Coreia do Sul, NG- Nigéria, PT- Portugal, SE- Suécia, SN- Senegal, TZ- Tanzânia, US- Estados Unidos da América, YE- Iêmen) e, finalmente, do número de acesso do isolado viral. A análise filogenética foi conduzida no programa *MEGA v 4.0* e a história evolutiva baseada no método filogenético *Neighbor-Joining* baseado no modelo 2 parâmetros de Kimura para o cálculo da matriz de distâncias genéticas (a escala indica 2% de diversidade genética). A robustez da inferência filogenética foi avaliada por *bootstrapping* de 1000 replicados, considerando-se representativos para genótipos valores iguais ou superiores a 70%. B) À direita da árvore filogenética encontram-se representados os resultados de *bootscanning* (janela 300 bp; passo 30 bp) por *SimPlot v 3.5.1* das 11 sequências consideradas assinaladas (com uma seta verde). As linhas horizontais pontilhadas indicam o *cut-off* de 70%, enquanto que as verticais sugerem a existência de um ponto de recombinação intra/inter-região (devido à natureza artificial da sequência, salienta-se que o ponto de recombinação inter-região é apenas virtual).

A análise da árvore filogenética revelou uma clara separação entre os grupos pertencentes a subtipos, sub-subtipos e formas recombinantes circulantes diferentes, sendo estes suportados por valores de *bootstrap* significativos (**Figura 3.13A**). Os resultados obtidos para as sequências em estudo são apresentados na **Tabela 3.7**.

Tabela 3.7. Resultados de genotipagem após construção da árvore filogenética com as sequências concatenadas (PR+RT+IN).

Genótipo provável	Sequência concatenada (n=38)
B	12 (31,6%)
F (F1)	1 (2,6%)
G ou G(14_BG)	13 (34,2%)
CRF02_AG	1 (2,6%)
Não atribuído*	11 (28,9%)

*Corresponde a sequências com posicionamento ambíguo na árvore filogenética ou demonstrando elevada divergência genética, seleccionadas para análise de recombinantes por *bootscanning*

A partir da análise da árvore filogenética construída com as 38 sequências concatenadas, 27 foram classificadas, desde logo, de um modo não ambíguo. Desta forma, treze sequências (PT_TP03, PT_TP13, PT_TP72, PT_TP98, PT_TX08, PT_TX12, PT_TX15, PT_TX23, PT_TX30, PT_TX32, PT_TX33, PT_TX45 e PT_TX48) foram subtipadas como G ou G(14_BG), visto formarem um grupo monofilético óbvio com sequências de referência G ou CRF14_BG, suportado por um valor de *bootstrap* elevado (93%) (**Figura 3.13A**). Foi verificado que as amostras PT_TX08 e PT_TX33 partilhavam o mesmo ancestral que a referência portuguesa G_PT_AY612637, enquanto que as restantes sequências em estudo não se agrupavam directamente com nenhuma referência, formando pequenos agrupamentos isolados. Verificou-se finalmente a presença de três sequências (PT_TX18, PT_TX40 e PT_TX51) com posicionamento ambíguo, mas claramente relacionadas com o genótipo G, seleccionadas para análise de *bootscanning*, para identificação de eventuais formas recombinantes.

No agrupamento definido pelas sequências de referência CRF02_AG, observou-se a inclusão de cinco sequências (PT_TP19, PT_TX13, PT_TX41 PT_TX46 e PT_TX52), formando um grupo monofilético muito evidente, suportado por um valor de *bootstrap* elevado (99%), embora seja perceptível a formação de três subgrupos

(**Figura 3.13A**). O primeiro é caracterizado pela presença de uma única sequência (PT_TX46), a qual é suportada por um ramo precocemente divergente. O segundo subgrupo é formado pela sequência PT_TP19 e duas sequências de referência, sendo suportado por um valor relativamente baixo de *bootstrap* (72%), ainda assim significativo. Assim, a sequência PT_TP19 foi subtipada como CRF02_AG. Por último, o terceiro subgrupo é constituído pelas sequências PT_TX13, PT_TX41 e PT_TX52, claramente relacionadas entre si. A posição ambígua na árvore filogenética do primeiro e terceiro subgrupos levantou dúvidas sobre a classificação das respectivas sequências, tendo-se por isso decido pela realização de uma análise de *bootscanning*.

A topologia do agrupamento definido pelas sequências PT_TP114, PT_TX16 e PT_TX47, suportado por um valor de *bootstrap* significativo (87%), evidenciando um relacionamento filogenético com os grupos anteriormente descritos, mas encontrando-se isolado na árvore (**Figura 3.13A**), foi factor determinante para a selecção das sequências em causa para a pesquisa de recombinantes.

Relativamente às restantes sequências em estudo, 13 (PT_TP02, PT_TP16, PT_TP24, PT_TP38, PT_TP71, PT_TP107, PT_TX21, PT_TX34, PT_TX50, PT_TX54, PT_TX59, PT_TX60 e PT_TX69) formaram um agrupamento robusto (*bootstrap* de 99%) com referências do subtipo B, embora uma delas (PT_TP107), devido ao seu posicionamento na árvore filogenética (**Figura 3.13A**), tenha sido seleccionada para análise de recombinantes. É de referir a presença de um subgrupo (aqui denominado **B^p**), constituído por 8 sequências, que não intercala com nenhuma das sequências de referência do subtipo B, ramificando em separado, embora com uma significância reduzida. De forma a averiguar se este conjunto de sequências forma um grupo distinto e consistente dentro do agrupamento B (eventualmente um sub-subtipo em circulação em Portugal), recorreu-se ao programa *Subtyping Distance Tool* (SUDI, disponível em <http://www.hiv.lanl.gov/content/sequence/SUDI/sudi.html>), para análise das distâncias genéticas (resultados não apresentados). Assim, foi possível concluir que:

as distâncias genéticas intra-**B^p** distribuem-se na gama das distâncias genéticas características dos subtipos já definidos;

o subtipo mais próximo é, como seria de esperar, o subtipo B;

as distâncias genéticas entre as sequências do subgrupo **B^P** e as sequências mais próximas geneticamente encontram-se numa gama de valores que indicam que **B^P** não constitui um sub-subtipo do subtipo B.

Finalmente, a sequência PT_AS646 foi subtipada como F, mais propriamente do sub-subtipo F1, pois forma um grupo monofilético óbvio, sustentado por um valor de *bootstrap* de 99%, com as sequências de referência daquele sub-subtipo (**Figura 3.13A**).

De seguida, e a fim de melhor caracterizar o conjunto das sequências nucleotídicas em estudo, procedeu-se à análise de *bootscanning* das onze sequências mencionadas nos parágrafos anteriores que revelaram posicionamentos ambíguos na árvore filogenética (**Figura 3.13A**). Um aspecto importante a salientar é que eventuais pontos de recombinação detectados entre PR/RT e RT/IN serão virtuais, devido à inexistência de continuidade entre as regiões, sendo desta forma impossível de os localizar com precisão, a existirem. Da análise efectuada verificou-se que oito das sequências (PT_TP107, PT_TP114, PT_TX16, PT_TX18, PT_TX40, PT_TX47, PT_TX51 e PT_TX52) revelavam natureza recombinante (**Figura 3.13B**). Relativamente às sequências PT_TX18 e PT_TX40, estas apresentavam uma estrutura mosaico análoga, $G(14_BG)^{PR}/G(14_BG)B^{RT}/G(14_BG)U^{IN}$ e $G(14_BG)^{PR}/B^{RT}/G(14_BG)^{IN}$, respectivamente. A sequência PT_TX51, por seu lado, possui homologia com o subtipo G(14_BG) nas extremidades PR e IN, aproximadamente, mas, entre as posições 675 e 1100 nucleótidos revelou perda do sinal filogenético para todos os subtipos testados (região do genoma U) (**Figura 3.13B**). Este facto explicará o grau de divergência genética demonstrado pela ramificação precoce desta sequência do agrupamento formado pelas referências G e G(14_BG), presentes na árvore filogenética. Por outro lado, foi observado que as sequências PT_TX16, PT_TX47 e PT_TP114, que formavam um grupo monofilético isolado de qualquer uma das referências presentes na árvore filogenética, suportado por um valor de *bootstrap* significativo (87%), apresentavam uma estrutura mosaico análoga, $B^{PR}/B^{RT}/G(14_BG)U^{IN}$, $B^{PR}/UB^{RT}/BG(14_BG)^{IN}$ e $G(14_BG)^{PR}/BUB^{RT}/G(14_BG)^{IN}$, respectivamente (**Figura 3.13B**). Quanto à sequência PT_TP107, esta revelou-se um recombinante com a estrutura $G(14_BG)^{PR}/B^{RT}/B^{IN}$ (**Figura 3.13B**), facto este que é indiciado pelo tamanho e posição do ramo que a suporta na árvore filogenética. Por

último, a sequência PT_TX52 demonstrou ser um recombinante com homologia quase total com o CRF02_AG, apresentando, contudo, a partir dos 1400 nucleótidos (região da IN), homologia com o G(14_BG), (**Figura 3.13B**), motivo pelo qual se agrupa com as sequências de referência da CRF02_AG.

Das onze sequências inicialmente seleccionadas para análise de *bootscanning*, três (PT_TX13, PT_TX41, PT_TX46) não revelaram uma natureza recombinante, possuindo homologia parcial com o CRF02_AG, acompanhada por regiões do genoma não genotipáveis (U), facto que estará na base do seu posicionamento ambíguo na árvore filogenética da **Figura 3.13A**, relativamente ao agrupamento definido pelas respectivas referências.

Ainda assim, e a fim de eliminar totalmente a possibilidade da divergência genética apresentada pelas sequências mencionadas ser devida ao fenómeno de hiperadenilação G→A (181), foi feita uma detecção de eventuais hipermutações induzidas pela APOBEC, implementada pelo programa *Hypermut v.2.0* (disponível em <http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html>). A pesquisa realizada não revelou indícios da ocorrência desse fenómeno, para qualquer uma das sequências analisadas (resultados não apresentados).

Em termos globais, pode afirmar-se que das 38 sequências concatenadas (PR+RT+IN) obtidas neste estudo, 31,6% (12/38) correspondiam a variantes genéticas do subtipo B e 34,2% (13/38) ao subtipo G ou G(14_BG). Das restantes sequências nucleotídicas, 34,2% (13/38) foram genotipadas como subtipos não-B/não-G, mais concretamente, 2,6% (1/38) do subtipo F (sub-subtipo F1), 10,5% (4/38) da CRF02_AG e 21,1% (8/38) de formas recombinantes únicas (URFs), este último um valor elevado (**Figura 3.14**).

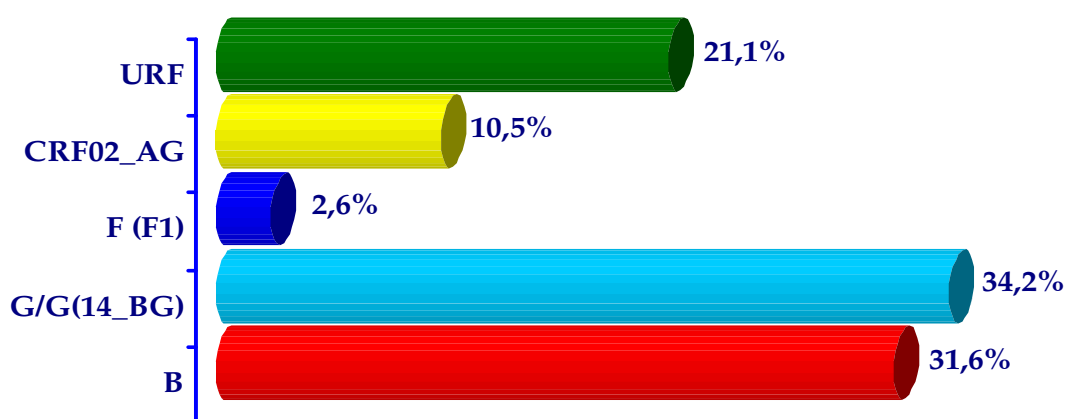


Figura 3.14. Representação gráfica da proporção relativa dos subtipos genéticos do HIV-1 obtidos a partir da análise filogenética de 38 sequências nucleotídicas concatenadas (PR+RT+IN).

3.6. Análise Bioinformática de Sequências de Aminoácidos

A informação disponível sobre a terapêutica anti-retroviral relativa aos 44 indivíduos infectados com vírus com pelo menos uma região amplificada, para os quais será pesquisada a presença de mutações de resistência aos anti-retrovirais, encontra-se resumida na **Figura 3.15**. Destes indivíduos, apenas quatro (9,1%) estavam sob a algum tipo de terapêutica anti-retroviral, à altura da colheita de sangue, o que contrasta com uma percentagem muito superior (17,6%) na amostra original (n=51) (**Figura 3.1**).

Como descrito no capítulo anterior, a análise de mutações associadas a diferentes níveis de resistência aos anti-retrovirais inibidores enzimáticos e de outros polimorfismos genéticos presentes nas sequências nucleotídicas correspondentes às regiões da PR, RT e IN do gene *pol* foi efectuada recorrendo-se ao programa *HIVdb*,

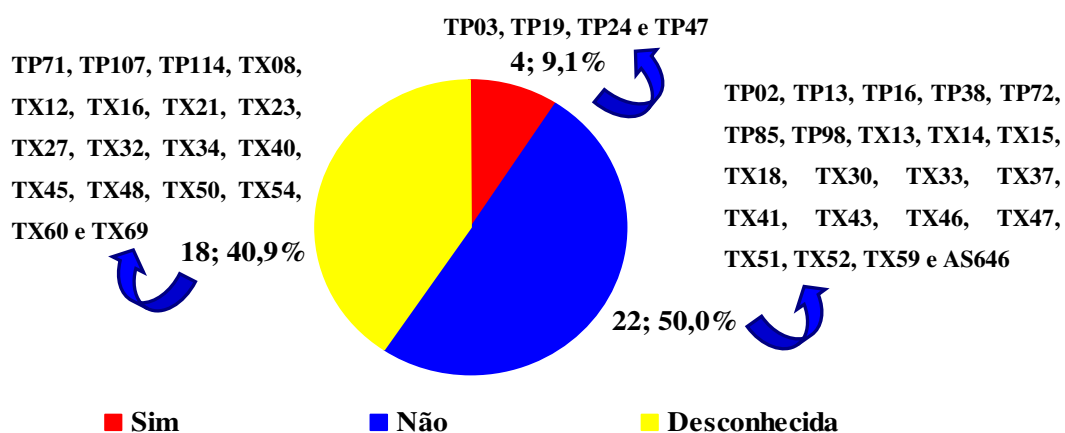


Figura 3.15. Representação gráfica dos resultados (em percentagem) da análise de distribuição relativa à terapêutica anti-retroviral para os 44 indivíduos infectados com pelo menos uma região amplificada do genoma do HIV-1.

implementado através do *Genotypic Resistance Interpretation Algorithm* da *Stanford University HIV Drug Resistance Database* (disponível em http://hivdb.stanford.edu/pages/algs/sierra_sequence.html), por comparação com uma sequência de aminoácidos consenso do subtipo B. Um relatório interpretativo foi obtido para cada sequência, mencionando as mutações associadas a resistência aos anti-retrovirais e outros polimorfismos genéticos. As sequências aminoacídicas estudadas correspondem à:

proteína PR completa (99 aminoácidos);

proteína RT incompleta (199/560¹ aminoácidos, posição 29-227, estando ausentes os 28 aminoácidos da extremidade N-terminal e os 333 aminoácidos que constituem a extremidade carboxílica terminal da RT);

proteína IN incompleta (271/288² aminoácidos, posição 1-271, estando ausentes os 17 aminoácidos da extremidade C-terminal da IN).

De realçar que não foram observadas deleções nem inserções em qualquer das sequências em estudo, o que resultou na ausência de mutações de *frameshift*. Contudo, foi observada a presença de um codão de terminação na posição 35 da sequência TX59_RT, embora apenas numa sub-população viral, devido à existência de *quasiespecies*.

3.6.1 Caracterização das mutações associadas a resistência

A pesquisa de mutações principais e acessórias associadas a resistência aos anti-retrovirais inibidores enzimáticos foi realizada após submissão ao programa *HIVdb* das 123 sequências nucleotídicas correspondentes às diferentes regiões codificantes do gene *pol*. O algoritmo utilizado faz também uma inferência sobre diferentes níveis de resistência (potencial baixo, baixo, intermédio e elevado), atribuindo uma pontuação a cada tipo de mutação encontrada, relativamente a cada uma das classes dos anti-retrovirais, incluindo oito inibidores da protease (PIs), sete inibidores nucleosídicos da transcriptase reversa (NRTIs), quatro inibidores não nucleosídicos da transcriptase reversa

¹ Tamanho total da subunidade maior (p66) da RT, relativamente a uma sequência de aminoácidos consenso do subtipo B.

² Tamanho total da IN, relativamente a uma sequência de aminoácidos consenso do subtipo B.

(NNRTIs) e dois inibidores da integrase (INIs; raltegravir, já licenciado pela FDA, e elvitegravir, em fase III de desenvolvimento).

3.6.1.1 Região da protease

Na região da PR não foram observadas mutações principais ou *major*. Quanto à presença de mutações acessórias, ou *minor*, foram encontradas quatro mutações distintas, que se encontram representadas na **Figura 3.16**.

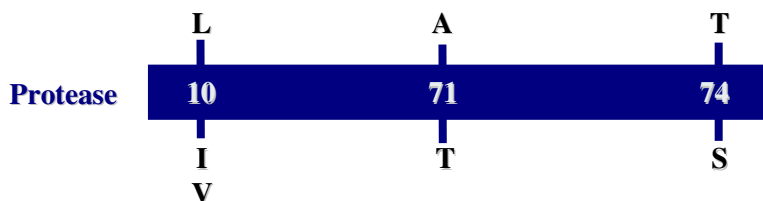


Figura 3.16. Representação esquemática das mutações de resistência acessórias presentes na região da protease do gene *pol* no grupo de 41 sequências em estudo. As letras em cima representam o aminoácido (código de uma letra, Anexo 5) existente na sequência consenso do subtipo B e a numeração identifica a respectiva posição. As letras em baixo correspondem ao aminoácido introduzido por efeito da mutação.

Em duas sequências do subtipo B foram encontradas as mutações L10V (PT_TX37) e A71T (PT_TX69), enquanto que L10I (PT_TX23 e PT_TX33), L10V (PT_TX46), A71T (PT_TP107 e PT_TX47) e T74S (PT_TX13, PT_TX41 e PT_TX52) encontram-se presentes em sequências de subtipos não-B (**Tabela 3.8**). Estas mutações foram observadas com maior frequência em sequências de subtipos não-B, comparativamente com as do subtipo B. Contudo, nenhuma se encontra associada de um modo significativo àqueles subtipos, conforme o resultado da análise estatística, utilizando o teste exacto de Fisher bilateral (22, disponível em <http://www.physics.csbsju.edu/stats/>), sobre tabelas de contingência 2x2, considerando um valor de significância de 5% (**Tabela 3.8**).

Tabela 3.8. Frequências absolutas das mutações de resistência aos inibidores da protease por subtipo (B vs. não-B), com indicação do valor de *p* associado à distribuição apresentada.

	Subtipo B (12 sequências)	Subtipos não-B (29 sequências)	<i>p</i>
L10I	0	2	1,000
L10V	1	1	0,505
A71T	1	2	1,000
T74S	0	3	0,543

Em termos de resposta à terapêutica, observou-se que 10,3% (3/29) das sequências não-B (PT_TX13, PT_TX41 e PT_TX52) apresentavam o perfil genotípico T74S, que confere resistência potencial de nível baixo ao inibidor da protease nelfinavir. A presença deste tipo de mutações não está associada a uma susceptibilidade diminuída aos anti-retrovirais, sendo apenas indicativa de eventual exposição prévia dos vírus que as possuem a uma determinada classe terapêutica (assim, estas mutações, por si só, não causam resistência aos anti-retrovirais) (205). Em conclusão, todas as mutações observadas estão associadas a um perfil fenotípico de susceptibilidade aos diferentes inibidores da PR (**Figura 3.17**).

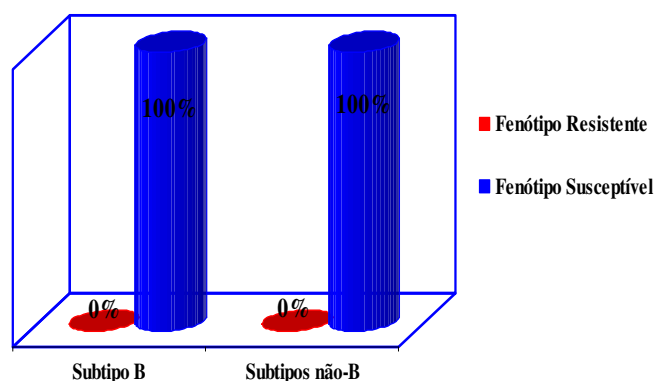


Figura 3.17. Representação gráfica dos perfis fenotípicos associados às mutações de resistência aos inibidores da protease encontradas em sequências nucleotídicas dos subtipos B e não-B.

3.6.1.2. Região da transcriptase reversa

Na região da RT não foram observadas mutações associadas a resistência aos NRTIs. Quanto aos NNRTIs, foram encontradas seis mutações distintas, que se encontram representadas na **Figura 3.18**.

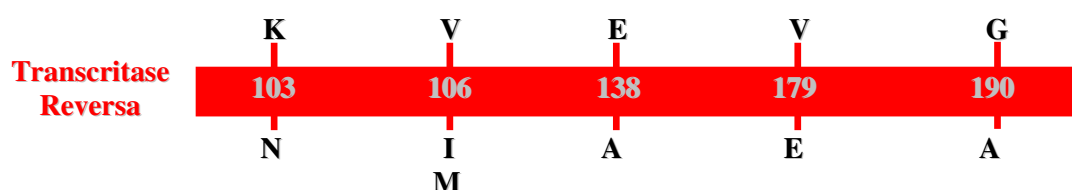


Figura 3.18. Representação esquemática das mutações de resistência aos inibidores não nucleosídicos da transcriptase reversa presentes na região da transcriptase reversa do gene *pol* no grupo de 41 sequências em estudo. As letras em cima representam o aminoácido (código de uma letra, Anexo 5) existente na sequência consenso do subtipo B e a numeração identifica a respectiva posição. As letras em baixo correspondem ao aminoácido introduzido por efeito da mutação.

As mutações K103N (PT_TP38, PT_TX16), V106I (PT_TP38, PT_TP107, PT_TP114, PT_TX16, TP_TX21 e PT_TX47), V106M (PT_TX21) e G190A (PT_TX16 e PT_TX21) encontram-se presentes em sequências do subtipo B. Nas

sequências de subtipos não-B observou-se V106I (PT_TX23), E138A (PT_TX13, PT_TX41 e PT_TX52) e V179E (PT_TP13). De salientar que a sequência PT_TX21 possuía duas mutações em simultâneo na mesma posição (V106I/M) devido à existência de *quasiespecies* virais. Por outro lado, e contrariamente ao observado para a região codificante da protease, três sequências (PT_TP38, PT_TX16 e PT_TX21) tinham mais do que uma mutação de resistência. Globalmente, e também em oposição ao caso anterior, as mutações foram observadas com maior frequência em subtipos B, comparativamente com as sequências dos subtipos não-B (**Tabela 3.9**). Foi constatado, conforme o resultado da análise estatística, utilizando o teste mencionado anteriormente, que a presença da mutação V106I se encontra associada de um modo significativo ao subtipo B ($p=0,014$) (**Tabela 3.9**).

Tabela 3.9. Frequências absolutas das mutações de resistência aos inibidores não nucleosídicos da transcriptase reversa por subtipo (B vs. não-B), com indicação do valor de p associado à distribuição apresentada.

	Subtipo B (17 sequências)	Subtipos não-B (24 sequências)	p
K103N	2	0	0,166
V106I	6	1	0,014
V106M	1	0	0,415
E138A	0	3	0,254
V179E	0	1	1,000
G190A	2	0	0,166

Em termos de resposta à terapêutica, constatou-se que 17,6% (3/17) das sequências B (PT_TP38, PT_TX16 e PT_TX21) apresentavam perfis genotípicos, embora distintos, associados a resistência aos NNRTIs, K103N/V106I, K103N/V106I/G190A e V106I/M/G190A, respectivamente. Estas mutações conferem resistência de nível elevado a delavirdine, efavirenz e nevirapine e de nível baixo ao etravirine. A presença *in vivo* deste tipo de mutações traduz uma quase ausência de resposta terapêutica e/ou que vírus com genótipo similar apresentaram *in vitro* os níveis mais elevados de resistência (205). Por outro lado, foi observado que uma das sequências não-B (PT_TP13) apresentava o perfil genotípico V179E, o qual confere resistência potencial de nível baixo aos diferentes NNRTIs, com a exceção do

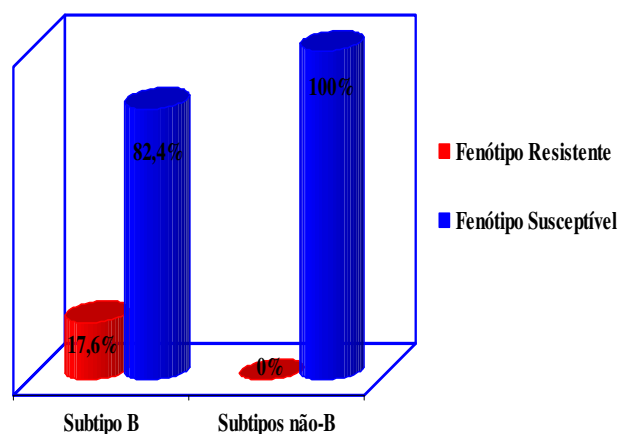


Figura 3.19. Representação gráfica dos perfis fenotípicos associados às mutações de resistência aos inibidores da transcriptase reversa (NNRTIs) encontradas em sequências nucleotídicas dos subtipos B e não-B.

etravirine. Por último, os perfis genotípicos V106I ou E138A, por si só, foram encontrados em três sequências do subtipo B (PT_TP107, PT_TP114 e PT_TX47) e quatro sequências dos subtipos não-B (PT_TX13, TX_23, PT_TX41 e PT_TX52). Estas mutações estão associadas a resistência aos NNRTIs, mas sozinhas estão relacionadas com um fenótipo de susceptibilidade.

Em conclusão, três das mutações encontradas (K103N, V106M e G190A) estão associadas a um perfil fenotípico de resistência a diferentes NNRTIs (**Figura 3.19**). Não foi observado qualquer tipo de associação entre a sua presença e os diferentes subtipos (B vs. não-B), conforme o resultado da análise estatística, utilizando o teste exacto de Fisher bilateral, nas condições já mencionadas ($p=0,064$).

3.6.1.3 Região da integrase

Na região da IN foram observadas uma mutação principal (S147G) e quatro mutações acessórias (T97A, G163R, I203M e S230N), representadas na **Figura 3.20**.

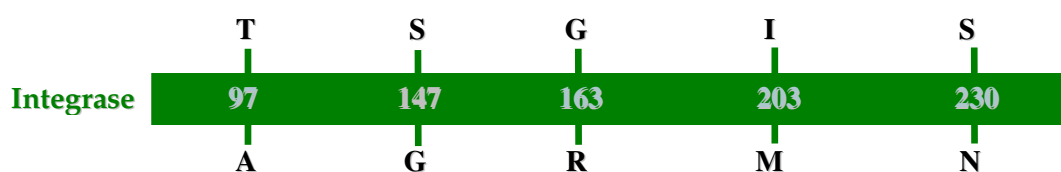


Figura 3.20. Representação esquemática das mutações de resistência primária e acessórias presentes na região da integrase do gene *pol*, no grupo de 41 sequências em estudo. As letras em cima representam o aminoácido (código de uma letra, Anexo 5) existente na sequência consenso do subtipo B e a numeração identifica a respectiva posição. As letras em baixo correspondem ao aminoácido introduzido por efeito da mutação.

A mutação principal S147G foi encontrada apenas uma vez, numa sequência de subtipo não-B (PT_TP72). Das quatro mutações acessórias, T97A e G163R foram observadas em sequências de subtipos não-B (PT_TP47 e PT_TX15, respectivamente), enquanto que I203M (PT_TP24, PT_TP107, PT_TX50, PT_TX54 e PT_TX69) e S230N (PT_TP38 e PT_TP71) em sequências B (**Tabela 3.10**). Estas mutações foram

observadas com muito maior frequência em sequências de subtipo B (7/13), comparativamente com as de subtipos não-B (3/28) (**Tabela 3.10**). Foi constatado que a presença da mutação acessória de resistência aos inibidores da integrase I203M se encontra associada de um modo altamente significativo ao subtipo B ($p=0,002$) (**Tabela 3.10**). Nenhuma das outras mutações revelou qualquer tipo de associação.

Tabela 3.10. Frequências absolutas das mutações de resistência aos inibidores da integrase por subtipo (B vs. não-B), com indicação do valor de p associado à distribuição apresentada.

	Subtipo B (13 sequências)	Subtipos não-B (28 sequências)	p
T97A	0	1	1,000
S147G	0	1	1,000
G163R	0	1	1,000
I203M	5	0	0,002
S230N	2	0	0,095

Relativamente ao potencial perfil fenotípico (**Figura 3.21**), observou-se que 3,6% (1/28) das sequências não-B (PT_TP72) apresentavam resistência de nível intermédio ao inibidor da integrase em desenvolvimento elvitegravir (conferido pela mutação

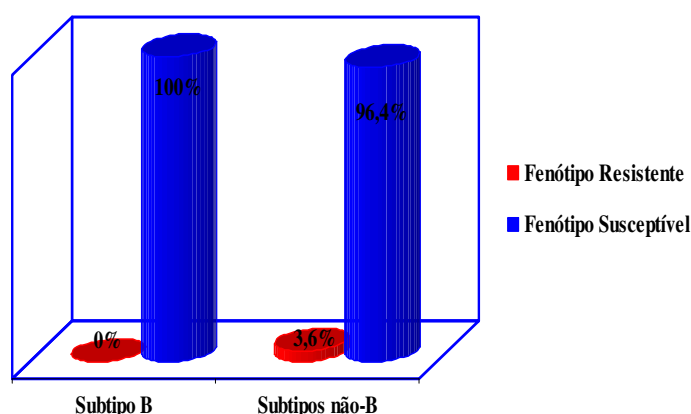


Figura 3.21. Representação gráfica dos perfis fenotípicos associados às mutações de resistência aos inibidores da integrase encontradas em sequências nucleotídicas dos subtipos B e não-B.

S147G). A presença deste genótipo sugere um grau de resistência ao fármaco já com um impacto significativo na sua eficácia (205). Por outro lado, todas as outras mutações encontradas, por si só, estão associadas a perfis fenotípicos de completa susceptibilidade, quer ao elvitegravir, quer ao único inibidor da integrase já licenciado, o raltegravir. Não

foi observado qualquer tipo de associação entre a presença de um potencial perfil fenotípico de resistência e os subtipos (B vs. não-B) ($p=1,000$).

No cômputo geral, foi constatado que a distribuição de mutações de resistência (perfil genotípico) variava entre 0 e 3 por indivíduo, relativamente às sequências PR, RT e IN (Figura 3.22). No entanto, a grande maioria dessas mutações, quando isoladamente, não induzem, necessariamente, o aparecimento de um perfil fenotípico de resistência à terapêutica anti-retroviral com inibidores enzimáticos. Isso veio a ser previsto para quatro dos indivíduos em estudo (9,1%, 4/44) e apenas para uma das classes de inibidores (da RT ou da IN) (Tabela 3.11).

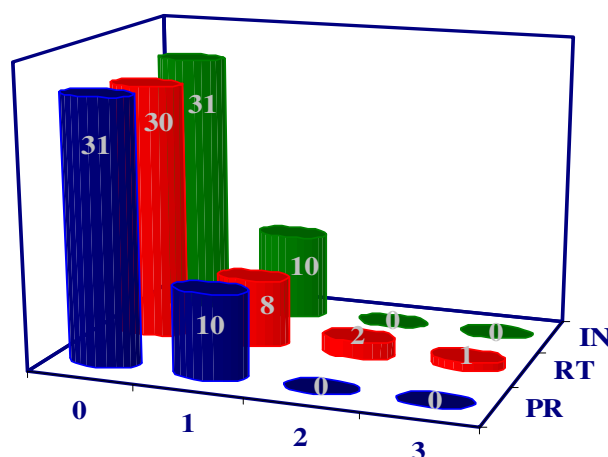


Figura 3.22. Representação gráfica da distribuição do número de mutações associadas a resistência, encontradas nas sequências da protease, transcriptase reversa e integrase do gene *pol*.

Tabela 3.11. Informação relevante sobre as sequências de HIV-1 que possuíam pelo menos uma mutação associada a um perfil fenotípico de resistência nas regiões do genoma em estudo (PR, RT e IN).

Sequência	Subtipo	Terapêutica anti-retroviral	Mutação	Perfil fenotípico
TP38_RT	B	Não	K103N	Resistência de nível elevado (DLV, EFV, NVP) [‡]
TP72_IN	G(14_BG)	Não	S147G	Resistência de nível intermédio (EVG) [§]
TX16_RT	B	Desconhecida	K103N e G190A	Resistência de nível elevado (DLV, EFV, NVP)
TX21_RT	B	Desconhecida	V106M e G190A	Resistência de nível elevado (DLV, EFV, NVP)

[‡]DLV- delavirdine, EFV- efavirenz, NVP - nevirapine; [§]EVG- elvitegravir.

3.6.2. Caracterização de outros polimorfismos genéticos

Por fim, foram ainda analisados os polimorfismos genéticos não directamente associados a resistência aos anti-retrovirais, presentes nas 123 sequências nucleotídicas correspondentes às diferentes regiões codificantes do gene *pol*. A categorização como

“outras mutações” pelo *Genotypic Resistance Interpretation Algorithm* da *Stanford University HIV Drug Resistance Database* foi baseada nos seguintes critérios:

serem consideradas mutações polimórficas (frequência superior a 1%, mesmo na ausência de pressão selectiva imposta pela terapêutica anti-retroviral);

possuírem um impacto nulo, ou quase nulo, na resistência fenotípica à terapêutica anti-retroviral;

representarem, muito frequentemente, a sequência consenso de subtipos não-B.

Uma listagem de todas as mutações encontradas nas três regiões codificantes do gene *pol* que obedecem aos critérios estabelecidos acima, e respectiva frequência absoluta, é apresentada no **Anexo 6**.

3.6.2.1. Região da protease

Das 99 posições aminoacídicas que constituem a PR, foi observado que 34,3% (34/99) apresentam, pelo menos, uma alteração relativamente à sequência consenso B. Como seria de esperar, constatou-se um elevado grau de conservação nos domínios funcionais da enzima. Nestes, apenas quatro polimorfismos genéticos (E21K, V56I, V82I e R87K) foram observados, no centro activo da enzima (posições aminoacídicas 21 a 33), *flap* (posições 45 a 56) e local de ligação do substrato (posições 78 a 88). É de salientar que a tríade catalítica de aminoácidos (D₂₅T₂₆G₂₇) presente no centro activo que caracteriza as proteases aspárticas dos lentívirus dos primatas se encontra absolutamente conservada.

A presença destes polimorfismos genéticos, não associados a resistência, foi observada com maior frequência em subtipos não-B, quando em comparação com as sequências do subtipo B. Foi verificado que quatro desses polimorfismos estão associados de um modo significativo ($p < 0,05$, teste exacto de Fisher bilateral) ao subtipo B e oito a subtipos não-B (**Figura 3.23, Anexo 6a**). Na maior parte (7/8, 87,5%), estes últimos foram considerados assinaturas (**Anexo 6d**) do subtipo G (**Figura 3.23–2**, a azul-escuro).

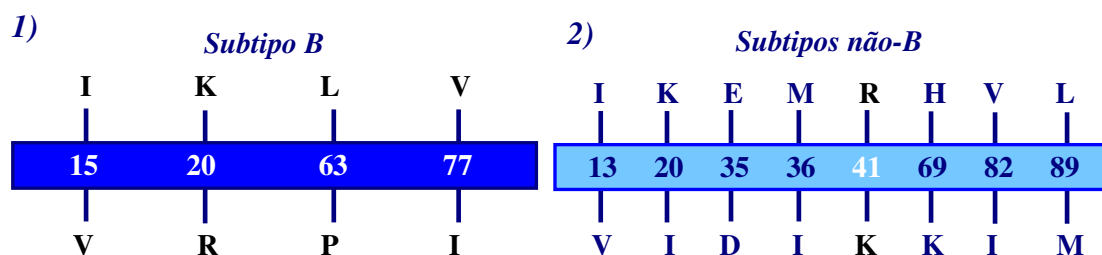


Figura 3.23. Representação esquemática dos polimorfismos genéticos não associados a resistência aos anti-retrovirais da região da protease do gene *pol*, associados, de um modo significativo, aos subtipos B (1) e não-B (2), presentes no grupo de 41 sequências em estudo. As letras em cima representam o aminoácido (código de uma letra, Anexo 5) existente na sequência consenso do subtipo B e a numeração identifica a respectiva posição. As letras em baixo correspondem ao aminoácido introduzido por efeito da mutação. Os polimorfismos genéticos considerados assinaturas do subtipo G estão assinalados a azul-escuro.

3.6.2.2 Região da transcriptase reversa

Relativamente à região da RT, foi observado que dos 199 aminoácidos em estudo, apenas 25,1% (50/199) correspondiam a posições polimórficas. Foi constatada a conservação absoluta de resíduos aminoacídicos essenciais para a manutenção da actividade catalítica da enzima (*e.g.* tríade catalítica D₁₁₀D₁₈₅D₁₈₆, motivo YMDD). Existem outros locais, não directamente relacionados com o centro activo (*e.g.* local de ligação aos dNTPs), que possuem outros resíduos que se encontram também absolutamente conservados, importantes para manter a funcionalidade da RT.

Globalmente, e em concordância com o caso anterior, este tipo de polimorfismos foi observado com maior frequência em subtipos não-B. Após análise estatística, verificou-se que quatro mutações distintas estavam associadas de um modo significativo ao subtipo B, enquanto que treze a subtipos não-B (**Figura 3.24, Anexo 6b**). Destas últimas, uma proporção elevada (11/13, 84,6%), foram consideradas assinaturas (**Anexo 6d**) do subtipo G (**Figura 3.24–2, a vermelho**), de acordo com os resultados dos testes estatísticos efectuados.

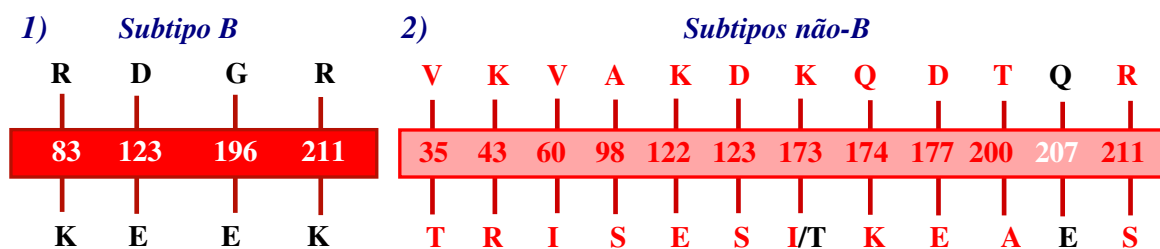


Figura 3.24. Representação esquemática dos polimorfismos genéticos não associados a resistência aos anti-retrovirais da região da transcriptase reversa do gene *pol*, associados, de um modo significativo, aos subtipos B (1) e não-B (2), presentes no grupo de 41 sequências em estudo. As letras em cima representam o aminoácido (código de uma letra, Anexo 5) existente na sequência consenso do subtipo B e a numeração identifica a respectiva posição. As letras em baixo correspondem ao aminoácido introduzido por efeito da mutação. Os polimorfismos genéticos considerados assinaturas do subtipo G estão assinalados a vermelho.

3.6.2.3 Região da integrase

Na região da IN constatou-se a existência de 29,9% (81/271) de posições polimórficas, após comparação das 41 sequências em estudo com a sequência consenso B. No geral, foi verificado que dos três domínios funcionais que constituem a enzima, o domínio catalítico (posições 50 a 212) revelou ser o mais conservado, com 24,5% (40/163) de posições polimórficas, comparativamente com os domínios N- (posições 1 a 49) e C-terminal (posições 213 a 271) que revelaram 34,7% (17/49) e 35,6% (21/59), respectivamente. Motivos importantes para a funcionalidade da enzima (domínio de multimerização HHCC e tríade catalítica D₆₄D₁₁₆E₁₅₂, entre outros) encontram-se absolutamente conservados. Tal como para as outras regiões codificantes estudadas, a presença dos polimorfismos foi observada com maior frequência em subtipos não-B, comparativamente com o subtipo B. Após análise estatística, constatou-se que cinco mutações distintas estavam associadas de um modo significativo ao subtipo B e doze a subtipos não-B ($p < 0,05$, teste exacto de Fisher bilateral) (**Figura 3.25, Anexo 6c**). Mais especificamente, os polimorfismos genéticos assinalados a verde na **Figura 3.25-2** foram considerados como assinaturas (**Anexo 6d**) do subtipo G (6/12, 50,0%).

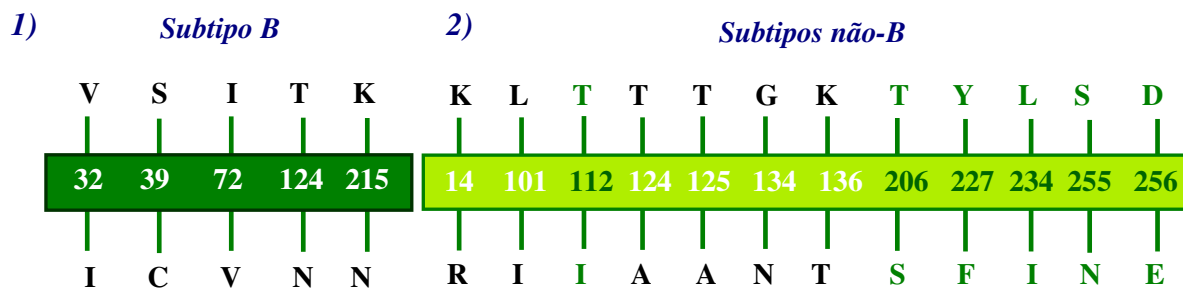


Figura 3.25. Representação esquemática dos polimorfismos genéticos não associados a resistência aos anti-retrovirais da região da integrase do gene *pol*, associados, de um modo significativo, aos subtipos B (1) e não-B (2), presentes no grupo de 41 sequências em estudo. As letras em cima representam o aminoácido (código de uma letra, Anexo 5) existente na sequência consenso do subtipo B e a numeração identifica a respectiva posição. As letras em baixo correspondem ao aminoácido introduzido por efeito da mutação. Os polimorfismos genéticos considerados assinaturas do subtipo G estão assinalados a verde.

4. Discussão e Conclusões

A epidemia HIV/SIDA em Portugal possui características específicas que as distinguem das epidemias da maior parte dos países europeus. Por um lado, os fluxos migratórios oriundos de África permitem que a infecção por subtipos não-B do HIV-1 se torne cada vez mais frequente (8, 60, 64, 65, 150, 151, 155). Por outro, um dos grupos de transmissão que mais contribui para a presente epidemia são os utilizadores de drogas por via endovenosa (IDUs). Esta constitui uma das formas mais comuns de exposição à infecção em Portugal, abrangendo 42,5% do total dos casos de HIV/SIDA notificados no país (31). Os IDUs têm desempenhado um papel importante em termos epidemiológicos constituindo uma fonte relevante de disseminação dos diferentes subtipos do HIV-1 e também de variantes virais resistentes aos diversos anti-retrovirais, potenciando ainda o aparecimento de novas formas genéticas.

A elevada diversidade genética é uma característica fundamental da biologia do HIV-1, ilustrando a sua estratégia de sobrevivência. Esta é caracterizada por uma rápida acumulação de mutações e pela ocorrência de fenómenos de recombinação entre diferentes estirpes virais (115, 216). Por outro lado, o tratamento prolongado com anti-retrovirais pode resultar na selecção de variantes genéticas do HIV-1 portadoras de mutações associadas a resistência à terapêutica actualmente disponível (115). Uma consequência evidente da elevada diversidade genética do HIV é o seu impacto no desenvolvimento de uma vacina eficaz. Menos óbvio, e bastante controverso, é, por exemplo, a sua importância na progressão da doença, transmissão vertical, resposta à terapêutica anti-retroviral e taxa de aquisição de mutações associadas a resistência aos diferentes inibidores (13). Daí o interesse da realização de estudos epidemiológicos moleculares para monitorização da prevalência dos diferentes subtipos e formas genéticas virais em circulação, procedendo-se, por exemplo, à genotipagem das sequências nucleotídicas obtidas nos testes genotípicos de resistência. Neste contexto, o presente estudo teve como principais objectivos:

caracterizar a diversidade genética das regiões codificantes da protease (PR), transcriptase reversa (RT) e integrase (IN) do gene *pol* em estirpes de HIV-1 circulantes numa população de IDUs na área da Grande Lisboa;

identificar os polimorfismos genéticos presentes nessas regiões do genoma, associados, ou não, a diferentes níveis de resistência aos anti-retrovirais inibidores enzimáticos já licenciados ou em vias de licenciamento;

estabelecer correlações entre os diferentes polimorfismos genéticos e as formas genéticas virais em circulação.

O gene *pol* do HIV-1 constituiu o objecto deste estudo, essencialmente, por duas razões. Este gene codifica três enzimas essenciais (PR, RT e IN) para o ciclo replicativo viral e, por outro lado, apresenta uma elevada conservação relativa, o que se reflecte numa das variabilidades mais baixas do genoma do HIV-1. Assim, os produtos do gene *pol* foram, desde cedo, considerados alvos principais da terapêutica anti-retroviral (18, 32, 204).

Para a realização deste tipo de estudo, a amostra biológica recomendada é o plasma, visto este ser representativo dos vírus livres em circulação, contendo, eventualmente, estirpes virais resistentes seleccionadas recentemente pela terapêutica anti-retroviral (195, 204). Foi sugerido que os vírus que circulam no plasma apresentam padrões de mutação diferentes dos detectados nas células mononucleadas de sangue periférico (PBMCs) e que este perfil de mutações possa preceder o existente no compartimento proviral (36). Outro aspecto a salientar é que os vírus que circulam no plasma representam a progénie de vírus que se encontram em replicação activa (195). Assim, a matriz genética inicial deste estudo foi o genoma de RNA viral, extraído de formas virais em circulação nos indivíduos infectados.

O sistema utilizado para a extracção do RNA viral a partir de 51 amostras de plasma de indivíduos toxicodependentes seropositivos para o HIV-1 foi o *INSTANT Virus RNA Kit* (Analytik Jena AG, Alemanha). Este sistema permite a extracção do RNA viral com elevada eficiência, sensibilidade e reprodutibilidade. Este passo é bastante crítico e tem que ser realizado com o máximo de cuidado, visto o RNA viral ser extremamente instável e facilmente degradado pelas RNases que se encontram presentes no meio ambiente. Posteriormente, para a realização do processo de retrotranscrição *in vitro* e primeira reacção de amplificação, recorreu-se ao sistema *Titan One Tube RT-PCR* (Roche Applied Science, Alemanha), o qual permite a realização de amplificações por RT-PCR de um modo rápido, sensível e reprodutível, a

partir de pequenas quantidades de RNA viral, com utilização de uma mistura de enzimas (a RT do AMV e as polimerases de DNA Taq e Tgo). Algumas das vantagens do sistema radicam nas características das enzimas, nomeadamente, a capacidade da RT do AMV superar facilmente a barreira das estruturas secundárias presentes no RNA viral, aumentando desta forma a eficiência da retrotranscrição, e a polimerase de DNA Tgo apresentar actividade de “correção de provas”, reduzindo a taxa de erro durante a reacção de amplificação. Por outro lado, o facto da retrotranscrição e primeira reacção de amplificação serem realizadas num único passo experimental representa uma menor probabilidade de contaminação. Ainda assim, as amostras para as quais não foi possível obter produtos de amplificação com esta estratégia experimental foram posteriormente utilizadas em experiências de RT-*nested* PCR convencional. O método de amplificação utilizado neste trabalho foi a *nested* PCR, cuja principal vantagem reside no aumento da sensibilidade e da especificidade da reacção de amplificação. Esta foi realizada recorrendo-se ao sistema *illustra™ puReTaq Ready-To-Go PCR Beads* (GE Healthcare, Inglaterra), o qual assegura níveis baixos de contaminação pela diminuição do número de passos de manipulação durante a preparação das misturas de reacção. Os protocolos de RT-*nested* PCR utilizados para a amplificação dos fragmentos da PR e da RT não necessitaram de um período inicial de optimização experimental, visto que as condições de amplificação (referidas no subcapítulo 2.6.1.3) já tinham sido optimizadas em estudos anteriores (207). Para a região da IN, no entanto, foi necessário fazer uma adaptação do procedimento experimental descrito por Braga (15), somente no que respeita à adição de um passo de retrotranscrição e à adaptação da temperatura de extensão durante a amplificação (descrito em 2.6.1.3). Contudo, antes de se iniciar o trabalho prático, as temperaturas de desnaturação T_M (do inglês, *melting temperature*), dos diferentes *primers* utilizados foram calculadas, com o objectivo de se confirmar as temperaturas de hibridação T_H (do inglês, *hybridization temperature*), aconselhadas nos protocolos em utilização. Neste cálculo, teve-se sempre em consideração a degeneração de alguns *primers* desenhados (PR1, RT137, RT3303, polOc_for, polO_for, polI_for e RT4), bem como o número de ligações GC a estabelecer. É de salientar que a hibridação dos *primers* na sequência alvo é absolutamente determinante para que as reacções de amplificação tenham sucesso. Assim, se a T_H for demasiado baixa ($T_H \ll T_M$), a probabilidade de ocorrência de hibridação dos *primers* em locais que não somente a

sequência alvo para a qual foram desenhados aumenta, amplificando-se produtos inespecíficos. Se, caso contrário, a T_H for demasiado elevada, há a probabilidade dos *primers* não hibridarem de todo. Uma estimativa correcta das temperaturas de hibridação dos *primers* é especialmente importante nas tentativas de amplificação de genomas com elevada variabilidade genética, como é caso do HIV-1. Por fim, de mencionar que a preparação das soluções-mãe das reacções de PCR foi sempre realizada em câmara de fluxo laminar e num espaço físico separado daquele onde foi adicionado o material genético matriz (adicionado sempre com pontas de micropipeta com filtro) e onde decorria a amplificação, a fim de evitar qualquer tipo de contaminação cruzada das amostras em estudo.

No presente estudo, o número total de produtos de amplificação obtidos foi de 126 (42 por região codificante), com origem em 44/51 indivíduos seleccionados (não foi possível obter qualquer produto de amplificação para sete destes indivíduos). Em 38/44 dos indivíduos com resultados positivos, foi possível obter amplicões para as três regiões em simultâneo. As possíveis razões para a não amplificação de algumas amostras estão abaixo mencionadas:

diferenças na sequência nucleotídica dos *primers* em relação aos respectivos locais de hibridação, apesar destas regiões serem consideradas relativamente bem conservadas;

degradação do RNA viral pelas RNases durante a sua extracção ou manuseamento experimental no decorrer do processo de retrotranscrição;

baixas cargas virais em indivíduos sob terapêutica anti-retroviral (a técnica de *nested* PCR utilizada pode não ter sensibilidade suficiente para detectar quantidades diminutas de material genético);

presença de estruturas secundárias no RNA viral, o que pode influenciar negativamente a eficiência da reacção de retrotranscrição, originando baixas quantidades de DNA complementar e, em consequência, quantidades muito reduzidas de DNA após amplificação por *nested* PCR.

Os produtos de amplificação obtidos que revelaram um fragmento específico correspondente às regiões da PR, RT e IN, de aproximadamente 460, 650 e 906 pb,

respectivamente, foram directamente sequenciados em ambas as cadeias. Para aqueles em que o fragmento pretendido surgiu acompanhado de outros fragmentos, resultantes de amplificações inespecíficas, havia quantidade reduzida de DNA ou, após sequenciação nucleotídica, eram obtidos cromatogramas difíceis de interpretar/editar, para os quais seria impossível obter uma sequência de qualidade, optava-se pela clonagem em vector plasmídico [*pGEM[®]-T Easy Vector* (Promega, EUA)]. A escolha deste sistema de clonagem deveu-se ao facto do vector acima referido, além de ser um plasmídeo de elevado número de cópias e de permitir a selecção rápida de clones bacterianos recombinantes baseada no sistema de α -complementação, apresentar em ambas as extremidades da região de clonagem múltiplas sequências repetidas de reconhecimento para endonucleases de restrição, permitindo o uso de uma única enzima (neste caso, a *EcoRI*) para o rastreio de clones positivos. Para além disso, a sua disponibilidade sob a forma linear, mantida pela existência de resíduos de timidina projectados nas extremidades 3', evitando assim a recircularização do próprio plasmídeo, facilita a clonagem directa dos produtos de PCR, uma vez que a polimerase de DNA Taq, durante a reacção de amplificação, adiciona um resíduo de desoxiadenosina em ambas as extremidades 3', independentemente da sequência do DNA matriz.

A selecção de clones bacterianos recombinantes foi realizada por métodos convencionais, por crescimento em meio LB contendo o antibiótico ampicilina, IPTG e X-gal. A presença de ampicilina permitiu a selecção de bactérias transformadas com *pGEM[®]-T Easy*, visto este vector plasmídico possuir o gene *bla* (do inglês, *beta-lactamase*), o qual confere resistência àquele antibiótico. Por outro lado, este sistema de clonagem permite a selecção simultânea das bactérias transformadas mas com os plasmídeos recombinantes. Na situação normal, o IPTG presente no meio vai induzir a expressão do gene *lacZ*, culminando na produção da enzima β -galactosidase, que, por sua vez, degrada o composto cromogénico X-gal, conferindo cor azul às colónias. A inserção dos fragmentos de DNA (460, 650 ou 906 pb) no vector plasmídico leva à interrupção do gene *lacZ*, cujo efeito se traduz na ausência de produção da enzima β -galactosidase, não ocorrendo por isso degradação do composto X-gal, originando colónias de cor branca. É de salientar que o tempo de incubação muito prolongado pode levar ao aparecimento de colónias brancas de dimensões menores (colónias satélite).

Este facto é explicado pela presença do gene *bla* que codifica para a produção de beta-lactamases, sendo que a presença destas enzimas no meio extracelular leva à inactivação da função bacteriostática da ampicilina, nas concentrações utilizadas, permitindo assim o crescimento destas colónias.

Após extracção, purificação e hidrólise com EcoRI dos plasmídeos recombinantes putativos, foram observados em gel de agarose padrões de restrição, entre os quais, alguns claramente indicadores da clonagem bem sucedida dos fragmentos de DNA pretendidos. O padrão de restrição mais comum era constituído por dois fragmentos de DNA, um correspondente ao vector plasmídico linearizado (aproximadamente 3000 pb) e outro, consoante a região estudada, correspondente ao fragmento de DNA clonado (para as regiões da RT e IN foi identificado, em alguns clones, um segundo local de reconhecimento para a enzima de restrição EcoRI). Um aspecto a realçar sobre o processo de clonagem é que esta metodologia pode fornecer informação adicional comparativamente à sequenciação directa de produtos obtidos por *nested* PCR. O estudo de um número significativo de clones diferentes, provenientes da mesma experiência de clonagem, permite o reconhecimento de variantes genéticas em circulação no mesmo indivíduo infectado, denominadas por *quasiespecies* virais. Estas variantes, com genomas muito idênticos entre si, mas com diferenças pontuais nas suas sequências nucleotídicas, aparecem devido a um conjunto de factores (referidos na secção 1.10), que são responsáveis, em última instância, pela existência de uma diversidade intra-individual elevada, limitando a eficácia da terapêutica anti-retroviral e dificultando a produção de vacinas. No entanto, a clonagem de todos os produtos amplificados, e, idealmente, o estudo de diversos clones por amostra, implicaria custos muito elevados, para além de ser um processo moroso, pelo que se optou pela sequenciação directa da maioria dos amplicões.

Em suma, de um total de 126 fragmentos de DNA enviados para sequenciação foram obtidas 123 sequências nucleotídicas de boa qualidade (97,6%). Apenas não foi possível realizar a edição de três sequências (2,4%), pelos motivos já descritos (ver secção 3.5.1), alguns dos quais poderão ter tido como causa problemas técnicos inerentes à própria sequenciação (por exemplo, por deficiente purificação do DNA a sequenciar) ou reflectirem, por outro lado, a extrema variabilidade genética apresentada pelo HIV-1.

De seguida, procedeu-se a uma pesquisa automática de homologia genética, recorrendo-se ao programa *BLAST*, apenas para confirmação preliminar da identidade das 123 sequências nucleotídicas e, posteriormente, à sua genotipagem automática recorrendo ao programa *REGA HIV-1 & 2 Automated Subtyping Tool*, v.2.0. Este programa permitiu caracterizar a diversidade genética da amostra em estudo, por intermédio de uma análise filogenética automatizada rápida, embora com limitações, combinada com métodos de *bootscanning* (somente para sequências superiores a 800 pb). Esta ferramenta bioinformática é mais exacta e credível do que outras disponíveis *online* que se baseiam na pesquisa de regiões de elevada similaridade local entre a sequência em questão e as sequências de HIV de genótipo conhecido depositadas em bases de dados específicas (50, 183). Esta análise automática revelou, desde logo, que a população viral em estudo possuía homologia genética maioritariamente com os subtipos B e G e as suas formas recombinantes.

O estudo mais aprofundado das sequências foi então iniciado através da construção de árvores filogenéticas, recorrendo a métodos não automatizados, que permitem a selecção, entre outros, dos algoritmos e das sequências de referência a utilizar, constituindo, por isso, métodos filogeneticamente muito mais fiáveis. O método e o modelo matemáticos utilizados para construção das árvores filogenéticas foram, respectivamente, *Neighbor-Joining* (184) e Kimura a dois parâmetros (108), para, cálculo de matrizes de distância. O primeiro é um dos métodos mais utilizados para análise filogenética, essencialmente devido à sua fiabilidade e eficiência, mesmo na análise de um conjunto grande de sequências nucleotídicas. Recorrendo ao cenário mais simples, duas sequências nucleotídicas diferentes, partilhando um ancestral, vão divergindo, ao longo do tempo, através da acumulação de mutações (substituições nucleotídicas). A necessidade de recorrer a modelos matemáticos, como o de Kimura a dois parâmetros (108), deve-se ao facto dos valores das distâncias genéticas estarem sempre enviesados, devido à possibilidade de ocorrência de substituições consecutivas nas mesmas posições, algumas, eventualmente, anuláveis. O modelo de substituição de nucleótidos utilizado estima, com base em certas premissas, o número de substituições nucleotídicas que realmente terão ocorrido, e, perante as diferenças existentes entre as sequências, calcula uma matriz de distâncias genéticas que serve como base para a construção da árvore filogenética. Este modelo distingue dois tipos de mutações

nucleotídicas, nomeadamente, transversões [substituição de uma purina (A e G) por uma pirimidina (C e T) ou vice-versa] e transições (substituição de purina por purina ou pirimidina por pirimidina), e assume que as últimas tendem a ocorrer mais frequentemente que as primeiras.

Numa primeira análise das árvores filogenéticas construídas para cada uma das três regiões codificantes (**Figuras 3.9A-3.11A**), constatou-se que a da região da PR foi a menos fiável, dada a existência de alguns agrupamentos monofiléticos suportados por valores de *bootstrap* abaixo do limite de confiança (70%). Muito provavelmente, este facto dever-se-á ao pequeno tamanho das sequências codificantes da PR (297 pb), consideravelmente menor do que o das outras regiões analisadas (RT 598 pb e IN 813 pb). Considerando ainda a relativa conservação das regiões codificantes para as enzimas virais, o segmento analisado originou pouca informação filogenética (“sinal filogenético” baixo). Assim, a classificação do genótipo das sequências em estudo para a região da PR foi realizada de um modo menos consistente, relativamente às outras regiões. Este facto corrobora os resultados de estudos realizados por Pasquier *et al* (157) e Gonzalez *et al* (83), os quais reportaram conclusões semelhantes. Desta forma, o tamanho do fragmento em análise, bem como o grau de variabilidade genética entre subtipos ou recombinantes geneticamente próximos, são dois factores importantes que condicionam a obtenção da informação necessária para uma análise filogenética consistente de um genoma (82). Para as regiões da RT e IN, a inferência filogenética realizada foi muito consistente, uma vez que as árvores filogenéticas construídas revelaram uma distinta separação entre os diferentes subtipos, com formação de agrupamentos monofiléticos suportados por valores de *bootstrap* significativos. No entanto, e a fim de aumentar ainda a entrada de informação nucleotídica em análise, procedeu-se à construção de uma árvore filogenética baseada numa sequência artificial, concatenada, resultante da simples justaposição das três sequências individuais PR, RT e IN (1707 nucleótidos, 38 sequências construídas). De salientar que na referida árvore, a ausência das sequências PT_TP47, PT_TP85, PT_TX14, PT_TX27, PT_TX37 e PT_TX43 deve-se ao insucesso de amplificação/edição de, pelo menos, uma das regiões codificantes.

De uma forma geral, os resultados obtidos da árvore construída com sequências concatenadas, como se esperava, foram, na sua maioria, concordantes com os da árvores

individuais, revelando uma clara separação entre os grupos de sequências pertencentes a subtipos, sub-subtipos e formas recombinantes circulantes diferentes, suportados por valores de *bootstrap* muito significativos (**Figura 3.13A**). A ocorrência de diferenças de classificação, quando comparados, entre si, os resultados obtidos na análise individual das três regiões codificantes, veio a ser justificada pela existência de um número significativo de formas recombinantes únicas, cuja estrutura foi completamente revelada por *bootscanning* (ver à frente).

Numa análise mais detalhada, as sequências genotipadas como G (n=13) apresentaram uma elevada proximidade filogenética, quer com formas G (puras) de referência em circulação em Portugal e Espanha, quer com as formas prototípicas da CRF14_BG, sugerindo que possam corresponder a esta forma recombinante (**Anexo 8**) ou às suas formas G progenitoras [daí serem sempre aqui designadas como formas G/G(14_BG)]. Uma informação adicional é que as amostras provenientes do INSA (prefixo PT_TX) tinham sido previamente genotipadas (por análise filogenética) para os genes *gag* (região desconhecida) e *env* (região da ansa V3) (**Anexo 7**). Considerando esses resultados, gentilmente cedidos pela Doutora Elizabeth Pádua (INSA), foi verificado que as amostras PT_TX30, PT_TX32 e PT_TX45 possuem identidade com a CRF14_BG, pois revelaram uma estrutura mosaico $G^{gag}/G^{pol}/B^{env}$, característica dessa forma recombinante circulante (**Anexo 8**). As sequências PT_TX08, PT_TX23, PT_TX33 e PT_TX48 revelaram, por seu lado, uma estrutura $G^{gag}/G^{pol}/G^{env}$, sugerindo serem formas G puras filogeneticamente relacionadas com as que deram origem à CRF14_BG. Para uma completa caracterização das sequências subtipadas neste trabalho como G/G(14_BG) seria necessário analisar o genoma completo dos respectivos vírus.

Segundo Esteves *et al* (65), as estirpes de HIV-1 do subtipo G circulantes em Portugal podem ter sido introduzidas há mais de duas décadas entre indivíduos portugueses toxicodependentes por via endovenosa, muito provavelmente a partir de indivíduos africanos, dada a sua circulação e elevada prevalência na África Central e Ocidental. Foi também demonstrado que no norte de Espanha (Galiza) se assiste à circulação de estirpes do subtipo G e de formas recombinantes circulantes suas derivadas (CRF14_BG), em associação a indivíduos toxicodependentes por via endovenosa, embora com uma menor prevalência, comparativamente com Portugal (221). A introdução do subtipo G no norte de Espanha poderá estar relacionada com a

actividade piscatória de longo curso e consequente contacto heterossexual com prostitutas africanas (53). Alguns autores, no entanto, sugerem que a forma recombinante B/G parental da CRF14_BG terá tido, muito provavelmente, um ancestral comum em Espanha (Galiza) e em Portugal (221). De salientar que em 2003 os recombinantes B/G representavam uma pequena minoria no nosso país (61, 65), embora, dada a proximidade geográfica, já houvessem indícios de que fizessem parte da mesma epidemia que os recombinantes circulantes no norte de Espanha (61). Dados mais recentes tornaram evidente que a forma recombinante B/G parental da CRF14_BG terá tido a sua origem em Portugal (~1990), logo após o início da epidemia de HIV/SIDA no nosso país, e daqui ter-se-á depois dispersado para Espanha (Galiza) (10, 164, 219). A co-circulação, em simultâneo, dos subtipos B e G no nosso país possibilitou o surgimento dos recombinantes B/G e a sua propagação para Espanha terá sido facilitada, muito provavelmente, pelo elevado número de indivíduos que atravessam a fronteira entre os dois países (10, 53), nomeadamente, indivíduos toxicodependentes por via endovenosa. Aparentemente, a CRF14_BG disseminou-se pela Europa, entre 1995 e 2005 (88), mas, segundo alguns autores (10), a sua prevalência terá diminuído desde então, não sendo conhecidas as razões para explicar tal facto. Por fim, a própria natureza do subtipo G tem sido objecto de aceso debate, nos últimos anos. Há estudos que apontam para que, contrariamente ao que fora anteriormente estabelecido, o subtipo G seja um recombinante, e não um subtipo puro, sendo a CRF02_AG a sua forma ancestral parcial (1). No entanto, dados mais recentes (245) vêm contrariar essa teoria, repondo a validade da classificação anterior.

Relativamente às sequências nucleotídicas genotipadas como B (n=12), verificou-se a existência de um sub-grupo que não intercalava nenhuma sequência de referência B, ramificando-se, em separado, constituído apenas por oito sequências portuguesas (PT_TP24, PT_TP38, PT_TP71, PT_TX21, PT_TX50, PT_TX54, PT_TX59 e PT_TX69), aqui denominado por B^p. De forma a determinar se este conjunto de sequências forma um grupo filogeneticamente consistente dentro do subtipo B, eventualmente um sub-subtipo (tal como descrito para os sub-subtipos A1-A4 e F1-F2), recorreu-se ao programa *Subtyping Distance Tool (SUDI)*. Baseado no cálculo e comparação de distâncias genéticas, este programa determina quando novas sequências representam novos subtipos ou sub-subtipos. Verificou-se que a distribuição dos valores

de distância genética resultantes da comparação entre as sequências do subgrupo B^p e as sequências de referência do subtipo B não se sobrepõe à gama típica de distribuição das distâncias genéticas entre sub-subtipos (dados não apresentados). Apesar deste resultado, e de um valor de *bootstrap* não significativo, a formação de um grupo monofilético separado (em todas as versões realizadas desta árvore filogenética) constituído somente por sequências de origem portuguesa sugere a possibilidade de um efeito fundador do subtipo B, através de uma única introdução ou introduções simultâneas de vírus geneticamente muito próximos, na população de IDUs portuguesas. Esta observação corrobora os resultados reportados por Esteves *et al.* (65), nomeadamente, a circulação na população de IDUs da Grande Lisboa de, pelo menos, duas subpopulações genéticas de vírus do subtipo B, reflectindo, muito provavelmente fenómenos específicos de efeito fundador, divergência e transmissão local. De salientar que os resultados de genotipagem realizados no INSA (**Anexo 7**) das amostras PT_TX21, PT_TX34, PT_TX50, PT_TX54, PT_TX59, PT_TX60 e PT_TX69, determinaram que estas são igualmente B para os genes *gag* e *env*, o que aponta para uma natureza não recombinante destas sequências. Por fim, digno de nota que em todas as árvores filogenéticas construídas, por simples comparação dos comprimentos dos ramos de suporte das sequências G e B, foi possível inferir uma maior divergência genética no grupo das sequências B, o que, assumindo taxas evolutivas semelhantes, sugere a sua circulação em Portugal há mais tempo que o subtipo G. Esta especulação vem corroborar o conhecimento da sua predominância inicial na Europa, sendo a infecção pelo subtipo G mais recente (64, 65).

Relativamente às sequências nucleotídicas genotipadas sem mais análises, pelo facto do seu posicionamento na árvore filogenética não se apresentar ambíguo, as sequências PT_AS646 e PT_TP19 foram classificadas respectivamente, como F, mais propriamente do sub-subtipo F1, e CRF02_AG (**Figura 3.13A**). Nas diferentes regiões do genoma analisadas, PT_AS646 aparece agrupada sempre com sequências F1, com valores de *bootstrap* muito significativos (**Figuras 3.9A, 3.10A, 3.11A**). Estudos anteriores (64, 65) revelaram que esta amostra, genotipada por HMA (do inglês, *heteroduplex multiplex assay*), é do subtipo F para o gene *env*, o que corrobora a hipótese de se tratar um subtipo puro F (F1). Quanto à sequência PT_TP19, o seu

posicionamento em todas as árvores filogenéticas construídas (**Figuras 3.9A, 3.10A, 3.11A, 3.13A**) não trouxe dificuldades relativamente à sua classificação.

No presente estudo, onze sequências nucleotídicas (28,9%) (PT_TP107, PT_TP114, PT_TX13, PT_TX16, PT_TX18, PT_TX40, PT_TX41, PT_TX46, PT_TX47, PT_TX51 e PT_TX52) apresentaram um posicionamento tal na árvore filogenética construída que dificultou a atribuição de genótipos, com grau elevado de certeza. Esta ambiguidade foi devida à presença de ramos divergentes (longos/graus de distância genética elevada) ou à não formação de grupos monofiléticos óbvios com os diferentes subtipos ou CRFs de referência do HIV-1, dispondo-se assim individualmente. A explicação para estas topologias pode radicar na ocorrência de fenómenos de recombinação intra-gene *pol*, na presença de estirpes virais geneticamente muito divergentes, relativamente às referências utilizadas, e ainda em fenómenos de hiperadenilação, caracterizados pela presença de hipermutação G→A. A fim de se testar esta última hipótese, recorreu-se ao programa *Hypermur* v.2.0 (181). Todavia, a pesquisa realizada não revelou indícios da ocorrência desse fenómeno.

A fim de investigar a presença de recombinantes intragénicos procedeu-se à análise de *bootscanning* das onze sequências concatenadas previamente mencionadas recorrendo-se ao programa *SimPlot* v.3.5.1 (120). Um aspecto importante a salientar é que eventuais pontos de recombinação detectados entre as regiões PR/RT e RT/IN serão virtuais, devido à ausência de continuidade entre as sequências das três regiões em estudo, sendo, desta forma, impossível de localizar com precisão os verdadeiros pontos de recombinação, a existirem. Da análise efectuada verificou-se que oito das sequências (PT_TP107, PT_TP114, PT_TX16, PT_TX18, PT_TX40, PT_TX47, PT_TX51 e PT_TX52) revelavam natureza recombinante e que as restantes três (PT_TX13, PT_TX41 e PT_TX46) apresentavam elevada homologia com CRF02_AG, excepto num pequeno troço na extremidade 5', que se revelou não genotipável (**Figura 3.13B**), explicando isto a topologia dos ramos que as suportam na árvore filogenética. Estas sequências foram então incluídas na CRF02_AG, mas como formas genéticas divergentes. Considerando a estrutura mosaico correspondente àquela forma recombinante circulante (**Anexo 8**) e a informação adicional fornecida para outros genes do genoma viral previamente estudados (A^{gag}/A^{env}) (**Anexo 7**), foi possível aumentar a fiabilidade daquela classificação.

Relativamente às sequências recombinantes, PT_TX18 e PT_TX40 apresentam uma estrutura mosaico análoga, respectivamente, $G(14_BG)^{PR}/G(14_BG)B^{RT}/G(14_BG)U^{IN}$ e $G(14_BG)^{PR}/B^{RT}/G(14_BG)^{IN}$ (**Figura 3.13B**), facto reflectido no seu agrupamento conjunto na árvore filogenética. Já a sequência PT_TX51 apresenta homologia com $G(14_BG)$, mas somente em ambas as extremidades (região central não genotipável) (**Figura 3.13B**). Este facto explicará a sua ramificação precoce do agrupamento formado pelas referências G e CRF14_BG, presentes na árvore filogenética (**Figura 3.13A**). Informação adicional para estas três sequências, relativa a regiões do genoma viral (*gag* e *env*) já estudadas, revelou uma estrutura B^{gag}/B^{env} para PT_TX18 e PT_TX40 e $B/J^{gag}/A^{env}$ para PT_TX51, o que confirma a sua essência recombinante única. Surge ainda a hipótese das duas primeiras representarem recombinantes de segunda geração, entre CRF14_BG e o subtipo B. Por outro lado, foi observado que as sequências PT_TX16, PT_TX47 e PT_TP114 formavam um agrupamento monofilético isolado de todas as referências presentes na árvore filogenética suportado por um valor de *bootstrap* significativo (87%). A ramificação precoce deste grupo, relativamente ao grande agrupamento G/G(14_BG) é explicada pela natureza recombinante intragénica das sequências que o constituem, apresentando, para além de alguma analogia entre si, regiões alternadas de genótipo B, $G(14_BG)$ e U [PT_TX16: $B^{PR}/B^{RT}/G(14_BG)U^{IN}$; PT_TX47: $B^{PR}/UB^{RT}/BG(14_BG)^{IN}$; PT_TP114: $G(14_BG)^{PR}/BUB^{RT}/G(14_BG)^{IN}$] (**Figura 3.13B**). Inicialmente foi colocada a hipótese destas sequências possuírem algum tipo de identidade com as formas recombinantes CRF20_BG, CRF23_BG ou CRF24_BG (**Anexo 8**), contudo esta hipótese não veio a ser confirmada (resultados não apresentados). Dado que PT_TP114 e PT_TX47 apresentam uma estrutura mosaico análoga no gene *pol*, não coincidente com nenhuma forma recombinante circulante descrita até ao momento, e que o valor de *bootstrap* que suporta o seu emparelhamento é altamente significativo (99%), é possível estarmos perante duas sequências com origem evolutiva comum. Para elucidar a verdadeira natureza destas sequências, nomeadamente se poderiam corresponder a uma nova CRF, seria necessária a sequenciação do seu genoma completo. A sequência PT_TX16, por seu lado, pode constituir uma forma recombinante única, provavelmente um recombinante de segunda geração. Como anteriormente, dados adicionais sobre outras regiões do genoma viral

previamente genotipadas (B^{gag}/B^{env}) corroboram essa hipótese. Quanto às restantes sequências, PT_TP107 revelou homologia inicial (~250 nucleótidos) com o G(14_BG), sendo o restante da sequência genotipado como B, razão pela qual se localiza na árvore filogenética com o grupo das sequências deste subtipo, mas numa posição precocemente divergente (**Figura 3.13A**). Já a sequência PT_TX52 demonstrou ser uma forma recombinante única, com homologia quase total para a CRF02_AG, apresentando, contudo, a partir dos 1400 nucleótidos (região da IN), uma maior identidade com o G(14_BG), motivo pelo qual se agrupa com as sequências de referência da primeira. A informação adicional previamente fornecida para outros genes do genoma viral (**Anexo 7**) indica que PT_TX52 apresenta uma estrutura A^{gag}/A^{env} , colocando-se a hipótese desta sequência ser uma URF de segunda geração.

De uma forma geral, os resultados obtidos revelaram um grau significativo de diversidade, sugerindo a existência de uma epidemia heterogénea relativamente às formas genéticas de HIV-1 circulantes na amostra de toxicodependentes estudada. A população viral revelou ser constituída maioritariamente pelos subtipos G (34,2%) e B (31,6%), havendo indícios claros da circulação de variantes de natureza recombinante *intra-pol* (21,1%), caracterizadas por estruturas mosaico maioritariamente homólogas daqueles dois subtipos e das formas recombinantes circulantes CRF02_AG e CRF14_BG. As restantes formas genéticas encontradas foram a CRF02_AG (10,5%) e o sub-subtipo F1 (2,6%). Estes resultados são genericamente concordantes com os dados provenientes de estudos prévios, nos quais foi constatada uma elevada diversidade genética de subtipos de HIV-1 circulantes em Portugal, para todas as regiões do genoma estudadas (9, 61, 152, 64, 65, 150, 155). Na Europa Ocidental, a epidemia de HIV-1 foi inicialmente dominada pelo subtipo B (21, 153), contudo, tem-se assistido em Portugal à co-circulação de estirpes não-B, em concordância com a situação descrita para outros países da Europa, como Espanha (221), França (45), Suíça (3) e Bélgica (89), por exemplo. No caso de Portugal, alguns autores associam as infecções não-B à numerosa comunidade imigrante, maioritariamente infectada por via heterossexual com diversos subtipos de HIV-1 adquiridos nos seus países de origem. Este facto relaciona-se não só com as ex-colónias portuguesas no continente africano, mas também com a Índia, o Brasil e, eventualmente, num futuro próximo, com os países da Europa de Leste, razão pela qual Portugal possui um perfil epidemiológico único,

distinto dos restantes países europeus (25, 152). No entanto, a homologia total ou parcial maioritariamente com os subtipos G e B e seus recombinantes (86,9% do total), apresentada pela maioria das sequências, reflectirá a predominância destes dois subtipos na população portuguesa de toxicodependentes por via endovenosa, pelo menos na área da Grande Lisboa. Este resultado corrobora ainda a ideia de Portugal poder estar a evoluir para um padrão epidemiológico único, onde os subtipos B, G e as suas formas recombinantes predominam (25, 64, 65, 152).

Posteriormente, foram caracterizadas as mutações associadas a diferentes níveis de resistência aos anti-retrovirais inibidores enzimáticos e outros polimorfismos genéticos presentes nas sequências nucleotídicas correspondentes às regiões da PR, RT e IN do gene *pol*. Esta pesquisa foi efectuada por recurso ao programa *HIVdb*, implementado através do *Genotypic Resistance Interpretation Algorithm* da *Stanford University HIV Drug Resistance Database*. A associação posterior das mutações encontradas a níveis de susceptibilidade fenotípica aos diferentes agentes anti-retrovirais foi realizada pelo mesmo algoritmo.

Os métodos genotípicos, como o utilizado neste trabalho, são utilizados com muito maior frequência para a detecção de mutações associadas a resistência aos anti-retrovirais, dado que são mais rápidos, mais baratos e de maior facilidade de realização, comparativamente com os métodos fenotípicos (estes, realizados *in vitro*, dão-nos uma medida relativa da capacidade de um isolado “selvagem” de HIV-1 replicar na presença de um determinado inibidor, em comparação com uma estirpe padrão susceptível) (13, 60, 114). Devido à existência de *quasiespecies*, é de notar que há populações virais minoritárias que persistem no plasma, não sendo, por vezes, detectáveis nos testes de resistência. Uma das grandes desvantagens dos métodos genotípicos consiste precisamente no facto de apenas serem detectadas estirpes resistentes que constituam pelo menos 20-30% do total da população viral (12). Por exemplo, foi reportado que a presença de estirpes virais minoritárias resistentes aos NNRTIs estava correlacionada com uma baixa resposta virológica, em indivíduos sob terapêutica de recurso incluindo estes inibidores (13). No entanto, uma investigação mais aprofundada é necessária para determinar o papel destas variantes minoritárias na aquisição de resistência aos anti-retrovirais, quer desta, quer de outras classes terapêuticas (13). Por fim, a interpretação da expressão fenotípica dos perfis genotípicos encontrados pelos algoritmos

automatizados está inteiramente dependente da realização prévia de estudos que correlacionem cada um dos padrões genotípicos de base com a resposta virológica *in vitro* ao tratamento. Assim, estes sistemas necessitam de uma actualização permanente das regras de interpretação, à medida que vão sendo descritos novos perfis de mutações de resistência e que vão sendo introduzidos novos fármacos no arsenal terapêutico.

Nos relatórios interpretativos obtidos para cada sequência no decurso deste trabalho, não foram reportadas deleções, nem inserções. Porém, foi detectada a presença de um codão de terminação na posição 35 da sequência TX59_RT, embora apenas numa subpopulação viral. A existência deste codão de terminação precoce, presente na sequência codificante da transcriptase reversa, faz questionar a capacidade replicativa desta subpopulação, após novo ciclo de infecção. Os dados epidemiológicos conhecidos (**Anexo 3**) indicam que o indivíduo em causa, à altura da colheita de sangue (2006), apresentava uma carga viral de 11000 cópias de RNA/ml, indicativo de replicação activa do vírus.

Das 41 sequências analisadas correspondentes à região da PR foram observadas quatro mutações acessórias (L10I/V, A71T, T74S). Neste estudo não foi observada a presença de mutações principais associadas à resistência aos inibidores da protease (PIs), ao contrário de estudos anteriores realizados em Portugal por Parreira *et al.* (154) e Palma *et al.* (152), nos quais essas mutações foram encontradas (*e.g.* D30N, N88S), contudo, numa frequência baixa. Em ambos os estudos, no entanto, foi observada uma elevada frequência de mutações acessórias. As mutações A71T e T74S foram encontradas em 7,3% (3/41) das sequências estudadas e as mutações L10I e L10V em 4,9% (2/41), para um total de dez sequências com alterações relativamente ao consenso. Segundo o algoritmo utilizado (205), os polimorfismos genéticos L10I/V ocorrem em 5 a 10% dos indivíduos nunca submetidos a terapêutica anti-retroviral (ou *naive*) e encontram-se associados a resistência, para a maioria dos PIs, apenas na presença de outras mutações. Relativamente a A71T, é um polimorfismo que ocorre basalmente em 2 a 4% dos indivíduos não tratados, aumentando a sua frequência em indivíduos sob terapêutica. Concordante com os dados anteriores, a mutação L10V ocorreu em 10,0% (2/20, PT_TX37 e PT_TX46) dos indivíduos *naive*. Contudo, nada se pode inferir neste trabalho acerca da frequência relativa em indivíduos não tratados do perfil genotípico L10I (presente em PT_TX23 e PT_TX33) e A71T (encontrado em PT_TP107,

PT_TX47 e PT_TX69), visto a informação relativa à terapêutica anti-retroviral ser desconhecida para três dos indivíduos em causa (**Figura 3.15**). Finalmente, o polimorfismo genético T74S é frequente em indivíduos *naive* infectados com estirpes virais do subtipo C (205), que não foram encontradas. Neste estudo, o perfil genotípico T74S ocorre em 15,0% (3/20, PT_TX13, PT_TX41 e PT_TX52) dos indivíduos não tratados, contudo apenas em sequências genotipadas como CRF02_AG (60,0% destas sequências apresentam esta mutação).

Em termos de potencial impacto fenotípico, apenas o perfil genotípico T74S parece ter alguma relevância, ainda que reduzida, quando cada uma das mutações encontradas é considerada *per se*, uma vez que se encontra associada a resistência potencial de nível baixo ao inibidor da protease nelfinavir. No entanto, e como se viu antes, a presença destas mutações não está associada a uma susceptibilidade diminuída aos anti-retrovirais, sendo apenas indicativa de uma eventual exposição prévia aos fármacos em questão (205). Curiosamente, nenhum dos três indivíduos (PT_TX13, PT_TX41, PT_TX52) reportou ter alguma vez sido submetido a terapia anti-retroviral. No geral, os resultados obtidos parecem corroborar o que está descrito na literatura, que a resistência aos PIs se desenvolve gradualmente através da acumulação de várias mutações, principais e acessórias (96), até se atingirem níveis clinicamente significativos na redução da susceptibilidade, caracterizando estes inibidores como de elevada barreira genética. Por outro lado, mesmo a maioria das mutações principais, e quando isoladamente, causam uma redução de duas a cinco vezes na susceptibilidade a um ou mais PIs, o que é geralmente insuficiente, só por si, para interferir com a sua actividade antiviral *in vivo* (13, 18).

Do total das sequências analisadas para a região da PR, constatou-se que 24,4% (10/41) continham uma das mutações acessórias descritas, sendo observada uma maior frequência em sequências dos subtipos não-B (27,6%, 8/29), comparativamente com as do subtipo B (16,7%, 2/12). Estes resultados corroboram estudos realizados por Vergne *et al.* (232) e Yebra *et al.* (242), nos quais se constatou uma maior frequência de mutações associadas a resistência aos PIs (maioritariamente acessórias) em subtipos não-B. Todavia, é de salientar que neste estudo não foi observado qualquer associação significativa entre a presença das referidas mutações e os subtipos (B vs. não-B) ($p=0,694$), confirmando os resultados de Monno *et al.* (137).

Relativamente à região da RT, nas 41 sequências analisadas foram encontradas seis mutações distintas (K103N, V106I/M, E138A, V179E e G190A), associadas a uma diminuição da susceptibilidade aos NNRTIs, não sendo observadas quaisquer mutações associadas a resistência aos NRTIs. Contrariamente, em estudos prévios, realizados por Palma *et al.* (152), Monno *et al.* (137) e Yebra *et al.* (242), foram observadas mutações de resistência associadas quer aos NRTIs, quer aos NNRTIs e numa frequência maior. Está referido na literatura (13) que estas últimas ocorrem normalmente entre as posições 98 a 108 e 179 a 190, o que foi constatado neste estudo.

No entanto, das mutações encontradas, apenas três estão associadas a perfis fenotípicos de resistência de nível elevado (K103N, V106M e G190A) a alguns NNRTIs (delavirdine, efavirenz e nevirapine). Dos três indivíduos infectados (TP38, TX16 e TX21), com vírus portadores de perfis genotípicos de resistência (K103N/V106I, K103N/V106I/G190A e V106I/M/G190A, respectivamente), apenas era conhecida informação sobre terapêutica anti-retroviral para o indivíduo TP38 (*naive*), o que valida a hipótese da ocorrência de transmissão contínua de estirpes virais resistentes aos anti-retrovirais nestes indivíduos utilizadores de drogas por via endovenosa (assumindo como correcta a informação fornecida pelos participantes no estudo aquando do preenchimento do inquérito epidemiológico básico). No cômputo global, de referir ainda que dos onze indivíduos infectados com vírus portadores de, pelo menos, uma das mutações encontradas, seis nunca tinham estado sob terapia anti-retroviral sugerindo, uma vez mais, a circulação destes vírus na população estudada. No entanto, na literatura está descrito (13) que as três mutações associadas a perfis fenotípicos de resistência de nível elevado acima referidas, ocorrendo ao nível do local hidrofóbico da RT, local onde os NNRTIs se ligam à enzima, aparecem normalmente associados à falha terapêutica com estes inibidores, traduzida pela supressão incompleta da replicação dos vírus. De salientar que a maior limitação desta classe de anti-retrovirais é a sua baixa barreira genética, uma vez que somente uma ou duas mutações pontuais neste local podem conferir, de imediato, um perfil fenotípico de resistência de nível elevado a um ou mais NNRTIs (13).

Por fim, no que respeita às restantes mutações encontradas (V106I, E138A e V179E), de acordo com os ensaios clínicos realizados no estudo DUET (236), as duas primeiras estão associadas a uma diminuição pouco significativa da resposta terapêutica

na presença de etravirine. Quanto a V179E, segundo o algoritmo utilizado (205), esta confere uma diminuição baixa da susceptibilidade ao delavirdine, efavirenz e nevirapine.

Em termos de distribuição, foi constatado que o número de mutações associadas a uma diminuição da susceptibilidade aos NNRTIs variava entre 1-3 por sequência nucleotídica (**Figura 3.22**). Observou-se ainda a presença de duas subpopulações virais distintas, ambas resistentes, para a sequência PT_TX21, relativamente à posição V106I/M, característico da existência de *quasiespecies* virais. A presença de uma degeneração (guanina ou adenina), na posição 271 desta sequência nucleotídica, leva à tradução dos codões ATA e ATG, os quais codificam, respectivamente, os aminoácidos isoleucina e metionina (**Anexo 5**). Como referido previamente, a presença de *quasiespecies* concorre para uma maior diversidade viral intra-individual e rápida evolução do HIV-1.

Das 41 sequências analisadas para a região da RT, foi observado que 26,8% (11/41) possuía, pelo menos, uma das mutações referidas. Contrariamente ao caso anterior (PR), constatou-se a existência destas mutações numa maior frequência em sequências do subtipo B (35,3%, 6/17), comparativamente com as dos subtipos não-B (20,8%, 5/24), embora esta diferença não se tenha revelado significativa ($p=0,476$). O resultado obtido não é concordante com Yebra *et al.* (242), os quais verificaram uma maior frequência de mutações associadas a resistência aos NNRTIs em sequências dos subtipos não-B. No presente estudo, a mutação V106I encontrava-se associada de forma significativa a sequências do subtipo B ($p=0,014$, **Tabela 3.9**), indicando ser um polimorfismo específico de subtipo ou assinatura (se descartarmos o efeito de selecção eventualmente exercido pela terapêutica anti-retroviral, uma vez que esse era um dado desconhecido para 5 dos 7 indivíduos infectados com vírus portadores da mutação V106I). Este resultado poderá ainda confirmar um estudo realizado por Asahchop *et al.* (5), no qual se constatou que esta mutação era preferencialmente seleccionada, em vírus do subtipo B (mas em indivíduos sob etravirine). Finalmente, de salientar que não foi presenciado qualquer tipo de associação significativa entre a presença de mutações relacionadas com um perfil fenotípico de resistência e os subtipos (B *vs.* não-B), o que confirma os dados de Monno *et al.* (137).

Quanto às 41 sequências da região da IN analisadas, foram encontradas cinco mutações distintas associadas a uma diminuição da susceptibilidade aos INIs, uma classificada como principal (S147G) e quatro acessórias (T97A, G163R, I203M e S230N). Demonstrou-se que 24,4% (10/41) das sequências continham uma das mutações descritas, sendo que cada uma delas continha apenas uma mutação (**Figura 3.22**). Neste estudo, apenas a sequência (PT_TP72), de subtipo não-B [mais propriamente, G(14_BG)], apresentava o perfil genotípico S147G. Não foram encontradas mutações nas posições 66, 92, 143, 148 e 155, as quais são referidas em diversos estudos (67, 72, 78) como possuindo um impacto considerável na diminuição da susceptibilidade ao raltegravir. A associação entre o aparecimento de mutações principais e a falha terapêutica a esse INI é mencionada em várias investigações realizadas (46, 67, 72). Por sua vez, as mutações acessórias T97A, G163R, I203M e S230N, observadas no presente trabalho, foram descritas também por Ceccherini-Silberstein *et al.* (26), Malet *et al.* (124), Fransen *et al.* (72) e Sichtig *et al.* (198) normalmente, em combinação com as mutações principais Y143R/C ou N155H, afectando o nível de susceptibilidade viral ao raltegravir.

De mencionar que as amostras de plasma com origem no INSA (referenciadas com prefixo TX) correspondem necessariamente a indivíduos que nunca foram sujeitos a esquemas terapêuticos incluindo o raltegravir, visto a colheita de sangue ter sido realizada em 2006 (nesse ano, o referido INI não tinha sido ainda licenciado pela FDA). Em estudos realizados por Ceccherini-Silberstein *et al.* (26), Garrido *et al.* (78) e Lataillade *et al.* (116) concluiu-se que as mutações associadas a resistência aos INIs, sobretudo as principais, se revelam ausentes, ou em frequência muito reduzida, em indivíduos *naive*. Deste modo, e logo à partida, haveria uma baixa probabilidade destes indivíduos estarem infectados com vírus portadores de um perfil de mutações associado a um nível de resistência elevado ao raltegravir. Não será então de estranhar que a mutação principal S147G, aqui encontrada uma única vez, esteja associada a uma amostra colhida no Centro de Atendimento a Toxicodependentes das Taipas (prefixo TP), cuja colheita foi realizada já em 2008. Mesmo assim, e em termos fenotípicos, esta mutação está associada a um perfil de resistência de nível intermédio apenas ao elvitegravir (INI ainda em fase de teste), tendo um efeito nulo ou quase nulo na susceptibilidade ao raltegravir (205).

Relativamente à distribuição das mutações associadas à diminuição da susceptibilidade aos INIs pelos genótipos das sequências nucleotídicas, observou-se que aquelas eram muito mais frequentemente encontradas no subtipo B (53,8%, 7/13) do que nas sequências não-B (10,7%, 3/28), diferença que veio a revelar-se muito significativa ($p=0,005$). Contrariamente, Eshleman *et al* (63) e Garrido *et al.* (78) reportaram que a frequência de mutações associadas a resistência é maior em subtipos não-B, enquanto Yerli *et al* (243) e Charpentier *et al.* (33) descreveram a mesma frequência de mutações entre os dois grupos de sequências. No entanto, como já referido, a única mutação principal (S147G) foi encontrada numa sequência de subtipo não-B. A análise estatística revelou ainda uma associação muito significativa entre a presença da mutação I203M e o genótipo B ($p=0,002$, **Tabela 3.10**), podendo aquela constituir uma assinatura das sequências de subtipo B circulantes nesta população de toxicodependentes portugueses. A existência da “assinatura B”, como a I203M, para a IN, é expectável face à presença de formas genéticas mais divergentes, evidenciada pela formação de um agrupamento distinto de sequências B portuguesas (**B^P**) ao nível da árvore filogenética concatenada (**Figura 3.13A**). A mesma associação (I203M vs. subtipo B) foi constatada num estudo de Garrido *et al.* (78), confirmando o resultado aqui obtido. Por fim, Van Baelen *et al.* (226) referiram que as mutações ditas acessórias, presentes na região codificante da integrase, tinham um efeito mínimo na susceptibilidade fenotípica ao raltegravir e ao elvitegravir (perfil fenotípico de susceptibilidade), quer em sequências do subtipo B, quer em subtipos não-B.

Adicionalmente, foram ainda analisados os polimorfismos genéticos não associados a resistência aos anti-retrovirais, presentes nas 123 sequências nucleotídicas correspondentes às três regiões codificantes do gene *pol*. O significado biológico destes polimorfismos mantém-se objecto de debate, não sendo ainda muito claro qual a sua relevância, nomeadamente em termos clínicos, relativamente aos diferentes subtipos do HIV-1. Independentemente destes polimorfismos apresentarem um impacto nulo, ou quase nulo, na resistência fenotípica à terapêutica anti-retroviral, a sua presença não deve ser ignorada, uma vez que podem ter um papel importante ao nível da *fitness* viral (por exemplo, pelo aumento da capacidade replicativa do vírus), mas também na evolução para um potencial perfil fenotípico resistente, em resposta à pressão selectiva imposta pelos fármacos (13). Por outro lado, estes polimorfismos genéticos podem estar

ainda associados de um modo significativo a subtipos virais específicos (assinaturas), podendo servir então como marcadores epidemiológicos (13, 95, 223).

No geral, e após comparação com uma sequência consenso do subtipo B, nas 99 (PR), 199 (RT) e 271 (IN) posições aminoacídicas estudadas foi observada uma taxa de variabilidade maior para a PR (34,3%), comparativamente com a RT (25,1%) e a IN (29,9%). Num estudo realizado por Vergne *et al.* (233) foi observado que as sequências da PR apresentavam uma taxa de variabilidade maior do que as da RT, em vírus de subtipos não-B que infectavam indivíduos *naive*. Por outro lado, foi também descrito que, no mesmo tipo de hospedeiros, a região mais conservada das três é a da IN, seguida da RT e PR (175). Assim, os resultados descritos neste trabalho confirmam, genericamente, os dados da literatura. Por fim, nas três regiões codificantes estudadas foi observada uma maior frequência destes polimorfismos genéticos em subtipos não-B, quando em comparação com o subtipo B, confirmando também resultados de estudos anteriores (78, 94, 137, 163, 175, 233). Foram ainda definidas diversas assinaturas, relativamente ao subtipo B, aos subtipos não-B, quando considerados em conjunto, e ao subtipo G (**Figuras 3.23-25**). O facto das sequências em estudo terem sido comparadas com uma sequência consenso B e de, ainda assim, terem sido descritas diversas assinaturas do subtipo B (n=13, no total), leva-nos a concluir que as formas genéticas B portuguesas são geneticamente divergentes comparativamente com as sequências B de referência, o que reforça os resultados reportados por Esteves *et al.* (65). Relativamente às assinaturas não-B, estas foram encontradas em maior número, encontrando-se significativamente associadas ao subtipo G, especialmente no caso da PR e da RT. Num estudo realizado por Kantor *et al.* (104), foi observado que, na região da RT, o polimorfismo A98S estava associado de um modo significativo ao subtipo G, o que veio a ser confirmado neste trabalho (**Figura 3.24**). Contudo, num outro estudo (169) constatou-se que a frequência deste polimorfismo era significativamente superior em sequências de origem portuguesa e espanhola, comparativamente com outros países, demonstrando uma forte associação entre a sua presença e a origem ibérica dos vírus ($p<0,0001$). Este facto será indiciador de que este polimorfismo genético, mais do que uma assinatura G, possa ser considerado como o aminoácido consenso das amostras daquele subtipo com origem em Portugal e Espanha. No cômputo geral, observou-se um

grau de conservação elevado nos domínios funcionais das três enzimas, tal como tinha sido antecipado.

Em conclusão, constatou-se uma elevada prevalência de subtipos B e G e das suas formas recombinantes na epidemia do HIV-1 na população estudada de toxicodependentes por via endovenosa, o que, a par de trabalhos previamente realizados (64, 65, 152, 153), reflectirá, ainda que parcialmente, a diversidade genética dos vírus circulantes na região da Grande Lisboa. A homologia total ou parcial com aqueles dois subtipos, apresentada para a grande maioria das sequências, espelha eventualmente bem a predominância dos subtipos B e G naquela população portuguesa. Por outro lado, os resultados aqui descritos poderão reforçar a ideia de Portugal estar a evoluir para um padrão epidemiológico único, onde os subtipos B e G e as suas formas recombinantes predominam. A hipótese da existência de diversas formas recombinantes únicas (de primeira e/ou de segunda geração) na população estudada comprova a teoria de que o processo de recombinação possa funcionar como mecanismo de aceleração da diversificação genética do HIV-1 e, consequentemente, da sua evolução (148).

Quanto às mutações associadas a resistência aos inibidores enzimáticos, com expressão fenotípica, estas foram encontradas em 9,1% (4/44) dos indivíduos infectados e somente para a RT (K103N, V106M e G190A) e para a IN (S147G). A presença de mutações que conferem um perfil fenotípico de resistência aos anti-retrovirais pode constituir uma desvantagem selectiva para o vírus. Porém, a presença de mutações acessórias e de outros polimorfismos genéticos, observados com elevada frequência na população viral estudada, podem influenciar a sua capacidade replicativa durante o tratamento, melhorando a *fitness* viral. Foi observado que metade dos indivíduos (50,0%, 11/22), nunca submetidos a terapêutica anti-retroviral estava infectada com vírus possuindo pelo menos uma mutação associada a resistência aos inibidores enzimáticos, indicando que a sua transmissão se encontra em curso na população de toxicodependentes estudada. Outro aspecto importante, é que a maior frequência, no cômputo geral, de polimorfismos genéticos (associados ou não a resistência aos fármacos) encontrada para as sequências não-B possa encontrar-se enviesada pelo facto da sua comparação ser realizada contra uma sequência de consenso do subtipo B (13, 91, 198, 232). Perante isto, parece pois importante a futura inclusão de estirpes virais não-B nos algoritmos automáticos de pesquisa de mutações de resistência ou mesmo

nos testes *in vitro* de susceptibilidade aos anti-retrovirais em desenvolvimento. Esta observação indica que, eventualmente, a abordagem terapêutica e as estratégias iniciais de tratamento para indivíduos infectados com vírus de subtipos não-B deveriam ser implementadas com base em algoritmos específicos. De realçar que estudos adicionais sobre avaliação genotípica e fenotípica da resistência aos anti-retrovirais realizados com diversas formas genéticas do HIV-1, sobretudo os subtipos não-B, actualmente em franca expansão geográfica, são necessários, de modo a determinar se o subtipo viral influencia a emergência de resistência natural aos diferentes inibidores e qual o verdadeiro impacto dessa influência na eficácia da terapêutica. Desta forma, será essencial proceder à implementação da monitorização clínica e laboratorial contínua da emergência de mutações associadas a resistência aos inibidores enzimáticos (e outros), quer em indivíduos seropositivos *naïve*, quer em indivíduos já sob terapêutica anti-retroviral.

5. Referências Bibliográficas

1. **Abecasis, A. B., P. Lemey, N. Vidal, T. de Oliveira, M. Peeters, R. Camacho, B. Shapiro, A. Rambaut, and A. M. Vandamme.** 2007. Recombination confounds the early evolutionary history of human immunodeficiency virus type 1: subtype G is a circulating recombinant form. *J. Virol.* **81**:8543-8551.
2. **Al-Mawsawi, L. Q., and N. Neamati.** 2007. Blocking interactions between HIV-1 integrase and cellular cofactors: an emerging anti-retroviral strategy *Trends Pharmacol. Sci.* **28**:526-535.
3. **Alaeus, A, T. Leitner, K. Lidman, and J. Albert.** 1997. Most HIV-1 genetic subtypes have entered Sweden. *AIDS* **11**:199-202.
4. **Arien, K. K., A. Abraha, M. E. Quinõnes-Mateu, L. Kestens, V. Guido, and J. Eric.** 2005. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *J. Virol.* **79**:8979-8990.
5. **Asahchop, E. L., C. Tremblay, B. G. Brenner, M. Oliveira¹, D. Moisi, T. d'Aquin Toni, B. Spira, and M. A. Wainberg.** 2010. *In vitro* selection of a novel etravirine-associated resistance mutation in B and non-B HIV-1 subtypes, abstract 552. *In 17th Conference on Retroviruses and Opportunistic Infections* Sao Francisco, California.
6. **Ayoub, A., S. Souquière, B. Njinku, P. M. Martin, M. C. Müller-Trutwin, P. Roques, F. Barré-Sinoussi, P. Maucière, F. Simon, and E. Nerrienet.** 2000. HIV-1 group N among HIV-1-seropositive individuals in Cameroon. *AIDS* **14**:2623-2625.
7. **Barré-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dautet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**:868-871.
8. **Bártolo, I., C. Rocha, J. Bartolomeu, A. Gama, R. Marcelino, M. Fonseca, A. Mendes, M. Epalanga, P. Cavaco Silva, and N. Taveira.** 2009. Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: New insights into the origins of the AIDS pandemic. *Infect. Genet. Evol.* **9**:672-682.
9. **Bártolo, I., H. Barroso, J. M. Pereira, K. Mansinho, M. Mafra, M. Martins, Maria H. Lourenço, R. Camacho, R. Antunes, S. Bento, S. Figueiredo, V. Bezerra, and N. Taveira.** 2000. Epidemiologia molecular da infecção pelo HIV em Portugal. *In* SIDAnet, Associação Lusófona (ed.), 1º Congresso Virtual: À descoberta de desafios partilhados na luta contra a SIDA.
10. **Bártolo, I., P. Borrego, A. B. Abecassis, F. McCutchan, and N. Taveira.** 2010. Origem e história epidemiológica da forma recombinante circulatória de VIH-1 CRF14_BG. III Congresso da CPLP VIH/SIDA*IST, Lisboa.

11. **Basavapathruni, A., and K. S. Anderson.** 2007. Reverse transcription of the HIV-1 pandemic. *FASEB J.* **21**:3795-808.
12. **Bennett, D. E., R. J. Camacho, D. Otelea, D. R. Kuritzkes, H. Fleury, M. Kiuchi, W. Heneine, R. Kantor, M. R. Jordan, J. M. Schapiro, A. M. Vandamme, P. Sandstrom, C. A. Boucher, D. Van De Vijver, S. Y. Rhee, T. F. Liu, D. Pillay, and R. W. Shafer.** 2009. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance. 2009 Update. *PLoS One* **4**:e4724.
13. **Benzie, A., A. Bertoli, M. Boffito, C. Booth, C. Boucher, F. Brun-Vézinet, R. Camacho, A. Cozzi-Lepri, M. Däumer, A. De Luca, D. Descamps, M. Fisher, C. Foster, A. M. Geretti, B. Killingley, H. Lyall, T. Lutz, N. Mackie, A. G. Marcelin, K. Metzner, G. Nebbia, A. Newell, C. F. Perno, E. Poveda, I. Reeves, C. Sabin, M. Schutten, T. Sing, V. Soriano, S. Staszewski, C. Theodore, and N. Van Maarseveen.** 2006. *Antiretroviral Resistance in Clinical Practice*. Coordinating ed., A. M. Geretti. Medscript Lda, London.
14. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**:1513-1523.
15. **Braga, J. P.** 2008. Polimorfismos genéticos naturais da região codificante da integrase do vírus da imunodeficiência humana (HIV-1) e resistência a agentes anti-retrovirais. Tese de Mestrado. Universidade Nova de Lisboa.
16. **Broglia, R. A., Y. Levy, and G. Tiana.** 2008. HIV-1 protease folding and the design of drugs which do not create resistance. *Curr. Opin. Struct. Biol.* **18**:60-66.
17. **Bukrinsky, M.** 2004. A hard way to the nucleus. *Mol. Med.* **10**:1-5.
18. **Buonaguro, L., M. L. Tornesello, and F. M. Buonaguro.** 2007. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol.* **81**:10209-10219.
19. **Burns, C. C., L. M. Gleason, A. Mozaffarian, C. Giachetti, J. K. Carr, and J. Overbaugh.** 2002. Sequence variability of the integrase protein from a diverse collection of HIV type 1 isolates representing several subtypes. *AIDS Res Hum. Retroviruses* **14**:1031-1041.
20. **Busschots, K., A. Voet, M. De Maeyer, J. C. Rain, S. Emiliani, R. Benarous, L. Desender, Z. Debyser, and F. Christ.** 2007. Identification of the LEDGF/p75 binding site in HIV-1 integrase. *J Mol. Biol.* **365**:1480-1492.
21. **Butler, I., F., I. Pandrea, P. A. Marx, and C. Apetrei.** 2007. HIV Genetic diversity: biological and public health consequences. *Curr. HIV Res.* **5**:23-45.
22. **Callegari-Jaques, S. M.** 2005. *Bioestatística: princípios e aplicações*. Artmed ed., Porto Alegre, Brasil.

23. **Cane, P. A.** 2009. New developments in HIV drug resistance. *J. Antimicrob. Chemother.* **64**:37-40.
24. **Carrasco L., and J. M. Almendral del Rio.** 2006. Retroviridae, p. 281-305. Hélice, Madrid, Espanha.
25. **Carvalho, A. P., P. Gomes, I. Diogo, M. F. Gonçalves, J. Cabanas, A. Abecasis, M. C. Lobo, and R. Camacho.** 2002. Molecular epidemiology of HIV-1 infection in Portugal: evolution for a unique pattern in Europe? *In* 8th European workshop on virus evolution and molecular epidemiology, Leuven, Belgium.
26. **Ceccherini-Silberstein, F., I. Malet, R. D'Arrigo, A. Antinori, A. G. Marcelin, and C. F. Perno.** 2009. Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev.* **11**:17-29.
27. **Centers Control Disease and Prevention (CDC).** 1982. Epidemiologic notes and reports persistent, generalized lymphadenopathy among homosexual males. *MMWR Morb. Mortal. Wkly. Rep.* **31**:249-251.
28. **Centers for Disease Control and Prevention (CDC).** 1981. Kaposi's sarcoma and *Pneumocystis pneumonia* among homosexual men-New York City and California. *MMWR Morb. Mortal. Wkly. Rep.* **30**:305-308.
29. **Centers for Disease Control and Prevention (CDC).** 1981. *Pneumocystis* -Los Angeles. *MMWR Morb. Mortal. Wkly. Rep.* **30**:250-252.
30. **Centers for Disease Control and Prevention (CDC).** 1982. Update on acquired immune deficiency syndrome (AIDS)-United States. *MMWR Morb. Mortal. Wkly. Rep.* **31**:507-508.
31. **Centro de Vigilância Epidemiológica das Doenças Transmissíveis, INSA.** 2009. Infecção VIH/SIDA: A Situação em Portugal. Lisboa, Portugal.
32. **Cerqueira, D. M., E. D. Ramalho, C. P. Oliveira, R. R. Silva, M. Franchini, M. S. Felipe, and C. R. Martins.** 2004. HIV-1 subtypes and mutations associated to antiretroviral drug resistance in human isolates from Central Brazil. *Braz. J. Microbiol.* **35**:187-192.
33. **Charpentier, C., M. Karmochkine, D. Laureillard, P. Tisserand, L. Belec, L. Weiss, A. Si-Mohamed and, C. Piketty.** 2008. Drug resistance profiles for the HIV integrase gene in patients failing raltegravir salvage therapy. *HIV Med.* **9**:765-770.
34. **Chen, Z., A., A. D. Luckay, L. Sodora, P. Telfer, P. Reed, A. Gettie, J. M. Kanu, J. M. Sadek, J. Yee, D. D. Ho, L. Zhang, and P. A. Marx.** 1997. Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J. Virol.* **71**:3953-3960.

35. **Chen, Z., P. Telfer, A. Gettie, P. Reed, L. Zhang, D. D. Ho, and P. A. Marx.** 1996. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J. Virol.* **70**:3617-3627.
36. **Chew, C. B., S. J. Potter, B. Wang, Y. M. Wang, C. O. Shaw, D. E. Dwyer, and N. K. Saksena.** 2005. Assessment of drug resistance mutations in plasma and peripheral blood mononuclear cells at different plasma viral loads in patients receiving HAART. *J. Clin. Virol.* **33**:206-216.
37. **Clavel, F., K. Mansinho, S. Chamaret, D. Guetard, V. Favier, J. Nina, M. O. Santos-Ferreira, J. L. Champalimaud, and L. Montagnier.** 1987. Human immunodeficiency virus type 2 infection associated with AIDS in West Africa. *N. Engl. J. Med.* **316**:1180-1185.
38. **Clavel, F., D. Guetard, F. Brun-Vezinet, S. Chamaret, M. A. Rey, M. O. Santos-Ferreira, A. G. Laurent, C. Dauguet, C. Katlama, C. Rouzioux, D. Klatzmann, J. L. Champalimaud, and L. Montagnier.** 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**:343-346.
39. **Clavel, F., M. Guyader, D. Guétard, M. Sallé, L. Montagnier, and M. Alizon.** 1986. Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature* **324**:691-695.
40. **Coffin, J., A. Haase, J. A. Levy, L. Montagnier, S. Oroszlan, N. Teich, H. Temin, K. Toyoshima, H. Varmus, and R. Weiss.** 1986. Human immunodeficiency viruses. *Science* **232**:697.
41. **Coffin, J., A. Haase, J. A. Levy, L. Montagnier, S. Oroszlan, N. Teich, H. Temin, K. Toyoshima, H. Varmus, P. Vogt, and R. A. Weiss.** 1986. What to call the AIDS virus? *Nature* **321**:10.
42. **Corbet, S., M. C. Muller-Trutwin, P. Versmisse, S. Delarue, A. Ayoub, J. Lewis, S. Brunak, P. Martin, F. Brun-Vezinet, F. Simon, F. Barre-Sinoussi, and P. Maucel.** 2000. *env* sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J. Virol.* **74**:529-534.
43. **Cornelissen, M., R. Van Den Burg, F. Zorgdrager, V. Lukashov, and J. Goudsmit.** 1997. *pol* gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D. *J. Virol.* **71**:6348-6358.

44. **Comunidade dos países de língua Portuguesa (CPLP) (ed).** 2008. Epidemia de VIH nos países de língua oficial portuguesa: Situação actual e perspectivas futuras rumo ao acesso universal à prevenção, tratamento e cuidados (disponível em http://www.aidsportugal.com/recursos/VIH_SIDA_em_CPLP.pdf).
45. **Couturier, E., F. Damond, P. Roques, H. Fleury, F. Barin, J. B. Brunet, and F. Brun-Vézinet, F. Simon.** 2000. HIV-1 diversity in France, 1996-1998. *AIDS* **14**:289-296.
46. **Croxtall, J. D., and S. J. Keam.** 2009. Raltegravir: A review of its use in the management of HIV infection in treatment-experienced patients. *Drugs* **69**:1059-1075.
47. **Dau, B., and M. Holodniy.** 2009. Novel targets for antiretroviral therapy: clinical progress to date. *Drugs* **69**:31-50.
48. **De Clercq, E.** 2008. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J. Antimicrob. Agents* **33**:307-320.
49. **De Clercq, E.** 2009. The history of antiretrovirals: key discoveries over the past 25 years. *Rev. Med. Virol.* **19**: 287-299.
50. **De Oliveira, T., K. Deforche, S. Cassol, M. Salminen, D. Paraskevis, C. Seebregts, J. Snoeck, E. Janse van Rensburg, A. M. Wensing, D. A. van de Vijver, C. A. Boucher, R. Camacho, and A. M. Vandamme.** 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* **21**:3797-3800.
51. **Deeks, S. G.** 2003. Treatment of antiretroviral-drug-resistant HIV-1 infection. *Lancet* **362**:2002-2011.
52. **Delelis, O., K. Carayon, A. Saïb, E. Deprez, and J. F. Mouscadet.** 2008. integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* **5**:114.
53. **Delgado, E., M. M. Thomson, M. L. Villahermosa, M. Sierra, A. Ocampo, C. Miralles, R. Rodríguez-Pérez, J. Diz-Aren, R. Ojea-de Castro, E. Losada, M. T. Cuevas, E. Vázquez-de Parga, R. Carmona, L. Pérez-Álvarez, L. Medrano, L. Cuevas, J. A. Taboada, and R. Nájera.** 2002. Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J. Acquir. Immune Defic. Syndr.* **29**:536-543.
54. **Delviks, K. A., and V. K. Pathak.** 1999. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching. *J. Virol.* **73**:7923-7932.

55. **Delwart, E. L., H. Pan, H. W. Sheppard, D. Wolpert, A. U. Neumann, B. Korber, and J. I. Mullins.**1997. Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. *J. Virol.* **10**:7498-7508.
56. **Department of Health and Human Services Panel on Antiretroviral Guidelines for Adults and Adolescents** (ed.). 2009. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. (disponível em <http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf>).
57. **Drew, W. L., L. Mintz, R. C. Miner, M. Sands, B. Ketterer.** 1981. Prevalence of cytomegalovirus infection in homosexual men. *J. Infect. Dis.* **143**:188-192.
58. **Dubey, S., Y. D. Satyanarayana, and H. Lavania.** 2007. Development of integrase inhibitors for treatment of AIDS: an overview. *Eur. J. Med Chem.* **42**:1159-1168.
59. **Dueweke, T. J., T. Pushkarskaya, S. M. Poppe, S. M. Swaney, J. Q. Zhao, I. S. Chen, M. Stevenson, and W. G. Tarpley.** 1993. A mutation in reverse transcriptase of bis (heteroaryl) piperazine-resistant human immunodeficiency virus type 1 that confers increased sensitivity to other nonnucleoside inhibitors. *Proc. Natl. Acad. Sci.* **90**:4713-4717.
60. **Duque, V.** 2005. Subtipos Genéticos e Resistência do HIV. Edições Minerva, **A. Melo-Silvestre, and J. G. Saraiva Da Cunha.** Coimbra, Portugal.
61. **Duque, V., A. Holguín, M. Silvestre, J. González-Lahoz, and V. Soriano.** 2003. Human immunodeficiency virus type 1 recombinant B/G subtypes circulating in Coimbra, Portugal. *Clin. Microbiol. Infect.* **9**:422-425.
62. **Dykes, C., K. Fox, A. Lloyd, M. Chiulli, E. Morse, and L. M. Demeter.** 2001. Impact of clinical reverse transcriptase sequences on the replication capacity of HIV-1 drug-resistant mutants. *Virology* **285**:193-203.
63. **Eshleman, S. H., S. E. Hudelson, P. Smith, J. Hackett, J. R. Holzmayer, P. Swanson, S. G. Devare, and N. Marlowe.** 2009. Analysis of pol integrase sequences in diverse HIV type 1 strains using a prototype genotyping assay. *AIDS Res. Hum. Retroviruses* **25**:343-345.
64. **Esteves, A., R. Parreira, T. Venenno, M. Franco, J. Piedade, J. Germano De Sousa, and W. F. Canas-Ferreira.** 2002. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes *AIDS Res. Hum. Retroviruses* **18**:313-325.
65. **Esteves, A., R. Parreira, J. Piedade, T. Venenno, M. Franco, J. Germano De Sousa, L. Patrício, P. Brum, A. Costa, and W. F. Canas-Ferreira.** 2003. Spreading of HIV-1 subtype G and *envB/gagG* recombinant strains among injecting drug users in Lisbon, Portugal. *AIDS Res. Hum. Retroviruses* **19**:511-517.

66. **Fang, G., B. Weiser, C. Kuiken, S. M. Philpott, S. Rowland-Jones, F. Plummer, J. Kimani, B. Shi, R. Kaul, J. Bwayo, O. Anzala, and H. Burger.** 2004. Recombination following superinfection by HIV-1. *AIDS* **18**:153-159.
67. **Ferns, R. B., S. Kirk, J. Bennett, I. Williams, S. Edwards, and D. Pillay.** 2009. The dynamics of appearance and disappearance of HIV-1 integrase mutations during and after withdrawal of raltegravir therapy. *AIDS* **23**:2159-2164.
68. **Flint, S. J., L. W. Enquist, R. M. Krug, V. R. Racaniello, and A. M. Skalka.** 2000. Principles of Virology: Molecular Biology, Pathogenesis, and Control. American Society for Microbiology, Washington, D.C.
69. **Food and Drug Administration (FDA) (ed.).** 2009. Isentress (raltegravir) indication extended for the treatment of HIV-1 infection in treatment-naïve patients. (disponível em <http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm171317.htm>).
70. **Forsman, A., and R. A. Weiss.** 2008. Why is HIV a pathogen? *Trends Microbiol.* **16**:555-560.
71. **Frankel, A. D., and J. A. Young.** 1998. HIV-1: Fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**:1-25.
72. **Fransen, S., S. Gupta, R. Danovich, D. Hazuda, M. Miller, M. Witmer, C. J. Petropoulos, and W. Huang.** 2009. Loss of raltegravir susceptibility of HIV-1 is conferred by multiple non-overlapping genetic pathways. *J. Virol.* **83**:11440-11446.
73. **Freed, E. O.** 2002. HIV-1 replication. *Somat. Cell Mol. Genet.* **6**:13-33.
74. **Furtado, M. R., D. S. Allaway, J. Phair, K. Kunstman, J. Stanton, C. Macken, A. Perelson, and A. Wolinsky.** 1999. Persistence of HIV-1 transcription in peripheral-blood mononuclear cells in patients receiving potent antiretroviral therapy. *N. Engl. J. Med.* **340**:1614-1622.
75. **Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai, B. White, P. Foster, and P. D. Markham.** 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**:500-503.
76. **Gao, F., E. Bailes, D. L. Robertson, Y. Chen, M. C. Rodenburg, S. F. Michael, L.B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature* **397**:436-441.
77. **Gao, Y., E. Paxinos, J. Galovich, R. Troyer, H. Baird, M. Abreha, C. Kityo, P. Mugenyi, C. Petropoulos, and E. J. Arts.** 2004. Characterization of a subtype D

- human immunodeficiency virus type 1 isolate that was obtained from an untreated individual and that is highly resistant to nonnucleoside reverse transcriptase inhibitors. *J. Virol.* **78**:5390-5401.
78. **Garrido, C., A. M. Geretti, N. Zahonero, C. Booth, A. Strang, V. Soriano, and C. De Mendoza.** 2010. Integrase variability and susceptibility to HIV integrase inhibitors: Impact of subtypes, antiretroviral experience and duration of HIV infection. *J. Antimicrob. Chemother.* **65**:320-326.
79. **Garrido, C., C. de Mendoza, and V. Soriano.** 2008. Resistencia a los inhibidores de la integrasa. *Enferm Infecc Microbiol Clin.* **12**:40-46.
80. **Geretti, A. M.** 2006. HIV-1 subtypes: epidemiology and significance for HIV management. *Curr. Opin. Infect. Dis.* **19**:1-7.
81. **Ghosh, M., P. S. Jacques, D. W. Rodgers, M. Ottman, J. L. Darlix, and S. F. Le Grice.** 1996. Alterations to the primer grip of p66 HIV-1 reverse transcriptase and their consequences for template-primer utilization. *Biochemistry* **35**:8553-8562.
82. **Gifford, R. J., T. de Oliveira, A. Rambaut, O. G. Pybus, D. Dunn, A. M. Vandamme, P. Kellam, D. Pillay, and on behalf of the UK collaborative group on HIV drug resistance.** 2007. Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. *J. Virol.* **80**:13050-13056.
83. **Gonzales, M. J., R. N. Machekano, and R. W. Shafer.** 2001. Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a Northern California clinic- based population. *J. Infect. Dis.* **184**: 998-1006.
84. **Gottlieb, M. S, R. Schroff, H. M. Schanker, J. D. Weisman, P, T, Fan, R. A. Wolf, and A. Saxon.** 1981. *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men. *N. Engl. J. Med.* **305**:1425-1431.
85. **Gutiérrez-Rivas, M, and L. Menéndez-Arias.** 2001. A mutation in the primer grip region of HIV-1 reverse transcriptase that confers reduced fidelity of DNA synthesis. *Nucleic Acids. Res.* **29**:4963-4972.
86. **Hahn, B. H., G. M. Shaw, K. M. De Cock, and P. M. Sharp.** 2000. AIDS as a zoonosis: scientific and public health implications. *Science* **287**: 607-614.
87. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**:95-98.

88. **Harris, B., I. Truchsess, H. Schatzl, S. Devare, and J. Hackett.** 2005. Genomic characterization of a novel HIV type 1 B/G intersubtype recombinant strain from an injecting drug user in Germany. *AIDS Res Hum Retroviruses*. **21**:654-660.
89. **Heyndrickx, L., W. Janssens, S. Coppens, K. Vereecken, B. Willems, K. Franssen, R. Colebunders, M. Vandenbruaene, and G. van der Groen.** 1998. HIV type 1 C2V3 *env* diversity among Belgian individuals. *AIDS Res. Hum. Retroviruses*. **14**:1291-1296.
90. **Hicks, C., and R. M. Gulick.** 2009. Raltegravir: the first HIV Type 1 integrase inhibitor. *Clin. Infect. Dis.* **48**:931-939.
91. **Hirsch, M. S., H. F. Gunthard, J. M. Schapiro, F. Brun-Vézinet, B. Clotet, S. M. Hammer, V. A. Johnson, D. R. Kuritzkes, J. W. Mellors, D. Pillay, P. G. Yeni, D. M. Jacobsen, and D. D. Richman.** 2008. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an international AIDS society–USA panel. *Clin. Infect. Dis.* **47**:266-285.
92. **Hirsch, V. M., R. A. Olmsted, M. Murphey-Corb, R. H. Purcell, and P. R. Johnson.** 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339**:389-392.
93. **Ho, D. D., and P. D. Bieniasz.** 2008. HIV-1 at 25. *Cell* **133**:561-565.
94. **Holguín, A., A. Alvarez, and V. Soriano.** 2002. High prevalence of HIV-1 subtype G and natural polymorphisms at the protease gene among HIV-infected immigrants in Madrid. *AIDS* **16**:1163-1170.
95. **Holguín, A., C. Suñe, F. Hamy, V. Soriano, and T. Klimkait.** 2006. Natural polymorphisms in the protease gene modulate the replicative capacity of non-B HIV-1 variants in the absence of drug pressure. *J. Clin. Virol.* **36**:264-271.
96. **Holguín, A., and V. Soriano.** 2002. Resistance to antiretroviral agents in individuals with HIV-1 non-B subtypes. *HIV Clin. Trials*. **3**:403-411.
97. **Hornak, V., and C. Simmerling.** 2007. Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov. Today*. **12**:132-138.
98. **Hou, T., W. A. McLaughlin, and W. Wang.** 2008. Evaluating the potency of HIV-1 protease drugs to combat resistance. *Proteins* **71**:1163-1174.
99. **Huet, T, R. Cheynier, A. Meyerhans, G. Roelants, and S. Wain-Hobson.** 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345**:356-359.
100. **Instituto di ecologia applicata (ed).** 1995-1999. African mammals databank. (disponível em <http://www.gisbau.uniroma1.it/amd/homespec.html>).

101. **Jablonski, J. A., E. Buratti, C. Stuardi, and M. Caputi.** 2008. The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. *J. Virol.* **82**:8038-8050.
102. **Jetz, A. E., H. Yu, G. J. Klarmann, Y. Ron, B. D. Preston, and J. P. Dougherty.** 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* **74**:1234-1240.
103. **Johnson, V. A., F. Brun-Vézinet, B. Clotet, H. F. Günthard, D. R. Kuritzkes, D. Pillay, J. M. Schapiro, and D. D. Richman.** 2008. Update of the drug resistance mutations in HIV-1: December 2008. *Top. HIV Med.* **16**:138-145.
104. **Kantor, R., D. A. Katzenstein, B. Efron, A. P. Carvalho, B. Wynhoven, P. Cane, J. Clarke, S. Sirivichayakul, M. A. Soares, J. Snoeck, C. Pillay, H. Rudich, R. Rodrigues, A. Holguin, K. Ariyoshi, M. Belen Bouzas, P. Cahn, W. Sugiura, V. Soriano, L. F. Brígido, Z. Grossman, L. Morris, A.-M. Vandamme, A. Tanuri, P. Phanuphak, J. N. Weber, D. Pillay, P. R. Harrigan, R. Camacho, J. M. Schapiro, and R. W. Shafer.** 2005. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: Results of a global collaboration. *PLoS Med.* **2**:e112.
105. **Katlama, C., and R. Murphy.** 2009. Emerging role of integrase inhibitors in the management of treatment-experienced patients with HIV infection. *Ther. Clin. Risk Manag.* **5**:331-340.
106. **Keele, B. F., F. V. Heuverswyn, Y. Li, E. Bailes, J. Takehisa, M. L. Santiago, F. Bibollet-Ruche, Y. Chen, L. V. Wain, F. Liegeois, S. Loul, E. M. Ngole, Y. Bienvenue, E. Delaporte, J. F. Y. Brookfield, P. M. Sharp, G. M. Shaw, M. Peeters, and B. H. Hahn.** 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**: 523-526.
107. **Kim, H. K., and E. S. Daar.** 2009. Newer antiretroviral agents and how to use them. *Curr. Infect. Dis. Rep.* **11**:327-334.
108. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of bases substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111-120.
109. **Kleiman, L., and S. Cen.** 2004. The tRNA^{Lys} packaging complex in HIV-1. *Int. J. Biochem. Cell Biol.* **36**:1776-1786.
110. **Kohl, N. E., E. A. Emini, W. A. Schleif, L. J. Davis, J. C. Heimbach, R. A. Dixon, E. M. Scolnick, and I. S. Sigal.** 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci.* **85**:4686-4690.
111. **Kolber, M. A., M. O. Saenz, T. J. Tanner, K. L. Arheart, S. Pahwa, and H. Liu.** 2008. Intensification of a suppressive HAART regimen increases CD4 counts and decreases CD8⁺ T-cell activation. *Clin. Immunol.* **126**:315-321.

112. **Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Bhattacharya.** 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**:1789-1796.
113. **Kosakovsky, S. L., and D. M. Smith.** 2009. Are all subtypes created equal? The effectiveness of antiretroviral therapy against non-subtype B HIV-1. *Clin. Infect. Dis.* **48**:1306-1309.
114. **Kozal, M. J.** 2009. Drug-resistant human immunodeficiency virus. *Clin. Microbiol. Infect.* **1**:69-73.
115. **Lal, R. B., S. Chakrabarti, and C. Yang.** 2005. Impact of genetic diversity of HIV-1 on diagnosis, antiretroviral therapy & vaccine development. *Indian J. Med. Res.* **121**:287-314.
116. **Lataillade, M., J. Chiarella, and M. J. Kozal.** 2007. Natural polymorphism of the HIV-1 integrase gene and mutations associated with integrase inhibitor resistance. *Antivir. Ther.* **12**:563-570.
117. **Levi, J. A., A. D Hoffman., S. M. Kramer, J. A. Landis, J. M. Shimabukuro, and L. S. Oshiro.** 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* **225**:840-842.
118. **Levin, A., J. Rosenbluh, Z. Hayouka, A. Friedler, and A. Loyter.** 2010. Integration of HIV-1 DNA is regulated by interplay between viral rev and cellular LEDGF/p75 proteins. *Mol. Med.* **16**:34-44.
119. **Liu, T. F., and R. W. Shafer.** 2006. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin. Infect. Dis.* **42**:1608-1808.
120. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**:152-160.
121. **London New Drugs Group on behalf of the HIV Drugs and Treatment subgroup of the London HIV Consortium (ed.).** 2009. London HIV Consortium (disponível em <http://www.nelm.nhs.uk/en/Original-search/?query=raltegravir+2009>).
122. **Low, A., N. Prada, M. Topper, F. Vaida, D. Castor, H. Mohri, D. Hazuda, M. Muesing, and M. Markowitz.** 2009. Natural polymorphisms of HIV-1 integrase and inherent susceptibilities to a panel of integrase inhibitors. *Antimicrob. Agents Chemother.* **53**:4275-4282.
123. **Makeyev, E. V., and D. H. Bamford.** 2004. Evolutionary Potential of an RNA Virus. *J. Virol.* **278**:2114-2120.

124. **Malet, I., O. Delelis, C. Soulie, M. Wirten, L. Tchertanov, P. Mottaz, G. Peytavin, C. Katlama, J. F. Mouscadet, V. Calvez, and A. G. Marcelin.** 2009. Quasispecies variant dynamics during emergence of resistance to raltegravir in HIV-1-infected patients. *J. Antimicrob. Chemother.* **63**:795-804.
125. **Malet, I., O. Delelis, M. A. Valantin, B. Montes, C. Soulie, M. Wirten, L. Tchertanov, G. Peytavin, J. Reynes, J. F. Mouscadet, C. Katlama, V. Calvez, and A. G. Marcelin.** 2008. Mutations associated with failure of Raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrob. Agents Chemother.* **52**:1351-1358.
126. **Mansky, L. M., and H. M. Temin.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **69**:5087-5094.
127. **Marchand, C., K. Maddali, M. Métifiot, and Y. Pommier.** 2009. HIV-1 IN inhibitors: 2010 update and perspectives. *Curr. Top. Med. Chem.* **9**:1016-1037.
128. **Markowitz, M., B. Y. Nguyen, E. Gotuzzo, F. Mendo, W. Ratanasuwan, C. Kovacs, G. Prada, J. O. Morales-Ramirez, C. S. Crumpacker, R. D. Isaacs, L. R. Gilde, H. Wan, M. D. Miller, L. A. Wenning, H. Teppler, and The Protocol 004 Part II Study Team.** 2007. Rapid and durable antiretroviral effect of the HIV-1 integrase inhibitor raltegravir as part of combination therapy in treatment-naïve patients with HIV-1 infection results of a 48-week controlled study. *J. Acquir. Immune. Defic. Syndr.* **46**:125-133.
129. **Marlink, R., P. Kanki, I. Thior, K. Travers, G. Eisen, T. Siby, I. Traore, C. C. Hsieh, M. C. Dia, E. H. Gueye, J. Hellinger, A. Guéye-Ndiaye, J. Sankalé, I. Mboup, and M. Essex.** 1994. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**:1587-1590.
130. **Martínez-Cajas, J. L., N. Pant-Pai, M. B. Klein, and M. A. Wainberg.** 2008. Role of Genetic Diversity amongst HIV-1 non-B subtypes in drug resistance: A systematic review of virologic and biochemical evidence. *AIDS Rev.* **10**:212-223.
131. **Marx, P. A.** 2005. Unsolved questions over the Origin of HIV and AIDS: Although HIV originated in nonhuman primates, we do not understand how it became a successful pathogen in its human host. *ASM News* **71**:15-20.
132. **Mathers, B., M., L. Degenhardt, B. Phillips, L. Wiessing, M. Hickman, S. A. A. Strathdee, A. Wodak, S. Panda, M. Tyndall, A. Toufik, and R. P. Mattick.** 2008. Global epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. *Lancet* **372**:1733-1745.
133. **Meadows, D. C., and J. Gervay-Hague.** 2006. Current developments in HIV chemotherapy. *ChemMedChem.* **1**:16-29.

134. **Menendez-Arias, L.** 2009. Molecular basis of human immunodeficiency virus drug resistance: An update. *Antiviral Res.* **85**:210-231.
135. **Miller, M. D., C. M. Hazuda, and F.D. Bushman.** 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **71**:5382-5390.
136. **Miyauchi, K., Y. Kim, O. Latinovic, V. Morozov, and G. B. Melikyan.** 2009. HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* **137**:433-444.
137. **Monno, L., L. Scudeller, G. Brindicci, A. Saracino, G. Punzi, A. Chirianni, A. Lagioia, N. Ladisa, S. Lo Caputo, and G. Angarano.** 2009. Genotypic analysis of the protease and reverse transcriptase of non-B HIV type 1 clinical isolates from naive and treated subjects. *Antiviral Res.* **83**:118-126.
138. **Montagnier, L., J. C. Chermann, F. Barré-Sinoussi, D. Klatzmann, S. Wain-Hobson, M. Alizon, F. Clavel, F. Brun-Vezinet, E. Vilmer, C. Rouzioux, W. Rozenbaum, and J. C. Gluckman.** 1984. Lymphadenopathy associated virus and its etiological role in AIDS. *Princess Takamatsu Symp.* **15**:319-331.
139. **Mougel, M., L. Houzet, and J. L. Darlix.** 2009. When is it time for reverse transcription to start and go? *Retrovirology* **6**:24.
140. **Moyle, G., J. Gatell, C. F. Perno, W. Ratanasuwan, M. Schechter and C. Tsoukas.** 2008. Potential for new antiretrovirals to address unmet needs in the management of HIV-1 infection. *AIDS Patient Care STDS* **22**:459-471.
141. **Mulky, A., and J. C. Kappes.** 2005. Analysis of human immunodeficiency virus type 1 reverse transcriptase subunit structure/function in the context of infectious virions and human target cells. *Antimicrob. Agents Chemother.* **49**:3762-3769.
142. **Mulky, A., B. C. Vu, J. A. Conway, S. H. Hughes, and J. C. Kappes.** 2007. Analysis of amino acids in the $\beta 7$ – $\beta 8$ loop of human immunodeficiency virus type 1 reverse transcriptase for their role in virus replication. *J. Mol. Biol.* **365**:1368-1378.
143. **Mulky, A., S. G. Sarafianos, Y. Jia, E. Arnold, and J. C. Kappes.** 2005. Identification of amino acid residues in the human immunodeficiency virus type-1 reverse transcriptase tryptophan-repeat motif that are required for subunit interaction using infectious virions. *J. Mol. Biol.* **349**:673-684.
144. **Nair, V., and G. Chi.** 2007. HIV integrase inhibitors as therapeutic agents in AIDS. *Rev. Med. Virol.* **17**:277-295.
145. **National Institute of Allergy and Infectious Diseases (NIAID).** 1995. The evidence that HIV causes AIDS. (disponível em <http://www3.niaid.nih.gov/topics/HIVAIDS/Understanding/howHIVCausesAIDS/HIVcausesAIDS.htm>).

146. **Nisole, S. N., and A. Saib.** 2004. Early steps of retrovirus replicative cycle. *Retrovirology* **1**:9.
147. **Novembre, F. J., M. Saucier, D. C. Anderson, S. A. Klumpp, S. P. O'neil, C. R. Brown Li, C. E. Hart, P. C. Guenther, R. Brent Swenson, and H. M. McClure.** 1997. Development of AIDS in a chimpanzee infected with human immunodeficiency virus type 1. *J. Virol.* **71**:4086-4091.
148. **Onafuwa-Nuga, A., and A. Telesnitsky.** 2009. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol. Mol. Biol. Rev.* **73**:451-480.
149. **Ortiz, A. M., and G. Silvestri.** 2009. Immunopathogenesis of AIDS. *Curr. Infect. Dis. Rep.* **11**:239-245.
150. **Pádua, E., R. Parreira, R. Tendeiro, B. Nunes, J. Castela, I. Soares, A. Mouzinho, E. Reis, and M. T. Paixão.** 2009. Potential impact of viral load and genetic makeup of HIV type 1 on mother-to-child transmission: Characterization of env-C2V3C3 and nef sequences. *AIDS Res. Hum. Retroviruses* **25**:1171-1177.
151. **Palma, A. C., A. B. Abecasis, J. Vercauteren, A. P. Carvalho, J. Cabanas, A. M. Vandamme, and R. J. Camacho.** 2010. Effect of human immunodeficiency virus type 1 protease inhibitor therapy and subtype on development of resistance in subtypes B and G. *Infect. Genet. Evol.* **10**:373-379.
152. **Palma, A. C., F. Araújo, V. Duque, F. Borges, M. T. Paixão, and R. Camacho.** 2007. Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect. Genet. Evol.* **7**:391-398.
153. **Paraskevis, D., O. Pybus, G. Magiorkinis, A. Hatzakis, A. M. Wensing, D. A. Van De Vijver, J. Albert, G. Angarano, B. Asjo, C. Balotta, E. Boeri, R. Camacho, M. L. Chaix, S. Coughlan, D. Costagliola, A. De Luca, C. De Mendoza, I. Derdelinckx, Z. Grossman, O. Hamouda, I. Hoepelman, A. Horban, K. Korn, C. Kuecherer, T. Leitner, C. Loveday, E. Macrae, I. Maljkovic, L. Meyer, C. Nielsen, E. L. Op De Coul, V. Ormaasen, L. Perrin, E. Puchhammer-Stockl, L. Ruiz, M. Salminen, J. C. Schmit, R. Schuurman, V. Soriano, and J. Stanczak.** 2009. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. *Retrovirology* **6**:49.
154. **Parreira, R., M. Santos, J. Piedade, and A. Esteves.** 2004. Natural polymorphism of HIV-1 subtype G protease and cleavage sites. *AIDS.* **18**:1345-1346.
155. **Parreira, R., E. Pádua, J. Piedade, T. Venenno, M. T. Paixão, and A. Esteves.** 2005. Genetic analysis of human immunodeficiency virus type 1 nef in Portugal:

- Subtyping, identification of mosaic genes, and amino acid sequence variability. *J. Med. Virol.* **77**:8-16.
156. **Parris, G. E.** 2007. How did the ancestral HIV-1 group M retrovirus get to Leopoldville from southeastern Cameroon? *Med. Hypotheses* **69**:1098-1101.
157. **Pasquier, C., N. Millot, R. Njouom, K. Sandres, M. Cazabat, J. Puel, and J. Izopet.** 2001. HIV-1 subtyping using phylogenetic analysis of *pol* gene sequences. *J. Virol. Methods* **94**:45-54.
158. **Peeters, M., C. Honoré, T. Huet, L. Bedjabaga, S. Ossari, P. Bussi, R. W. Cooper, and E. Delaporte.** 1989. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* **3**:625-630.
159. **Peeters, M, K. Fransen, E. Delaporte, M. Van den Haesevelde, G. M. Gershy-Damet, L. Kestens, G. van der Groen, and P. Piot.** 1992. Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. *AIDS* **6**:447-451.
160. **Peeters, M.** 2000. Recombinant HIV sequences: Their role in the global epidemic, p. 39-54. *In* C. Kuiken, B. Foley, B. Hahn, P. Marx, F. Mccutchan, J. W. Mellors, S. Wolinsky and B. Korber (ed.), *HIV Sequence Compendium 2000*, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico.
161. **Peeters, M., A. Gueye, S. Mboup, F. Bibollet-Ruche, E. Ekaza, C. Mulanga, R. Ouedrago, R. Gandji, P. Mpele, G. DibangaII, B. Koumare, M. Saidou, E. Esu-Williams, J. P. Lombart, W. Badombena, N Luo, M. V. Haesevelde, and E. Delaporte.** 1997. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **11**:493-498.
162. **Perelson, A. S., A. U. Neumann, M. Markowitz, M. J. Leonard, D. D. Ho.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**:1582-1586.
163. **Pérez-Alvarez, L, R. Carmona, M. Muñoz, E. Delgado, M. M. Thomson, G. Contreras, J. D. Pedreira, R. Rodríguez Real, E. Vázquez de Parga, L. Medrano, J. A. Taboada, Nájera R, and Spanish Group for Antiretroviral Resistance Studies in Galicia.** 2003. High incidence of non-B and recombinant HIV-1 strains in newly diagnosed patients in Galicia, Spain: study of genotypic resistance. *Antivir Ther.* **8**:355-360.
164. **Pérez-Alvarez, L., M. Munoz, E. Delgado, C. Miralles, A. Ocampo, V. García, M. Thomson, G. Contreras, R. Nájera, and The Spanish Group for Antiretroviral Resistance Studies in Galicia.** 2006. Isolation and biological characterization of HIV-1 BG intersubtype recombinants and other genetic forms circulating in Galicia, Spain. *J. Med. Virol.* **78**:1520-1528.

165. **Plantier, J. C., M. Leoz, J. E. Dickerson, F. De Oliveira, F. Cordonnier, V. Lemée, F. Damond, D. L. Robertson, and F. Simon.** 2009. A new human immunodeficiency virus derived from gorillas. *Nat Med.* **15**:871-872
166. **Poeschla, E. M.** 2008. Integrase, LEDGF/p75 and HIV replication. *Cell Mol. Life Sci.* **65**:1403-1424.
167. **Pollard, V. W.** 1998. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52**:491-532.
168. **Popovic M, M. G. Sarngadharan, E. Read, and R. C. Gallo.** 1984. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* **224**:497-500.
169. **Preiser, W, J. F. Drexler, and C. Drosten.** 2006. HIV-1 viral load assays for resource-limited settings: clades matter. *PLoS Med.* **3**:e538.
170. **Quercia, R., E. Dam, D. Perez-Bercoff, and F. Clavel.** 2009. Selective advantage profile of human immunodeficiency virus type 1 integrase mutants explain *in vivo* evolution of raltegravir resistance genotypes. *J. Virol.* **83**:10245-10249.
171. **Quinões-Mateu, M. E., D. M. Moore-Dudley, O. Jegede, J. Weber, and E. J. Arts.** 2008. Viral drug resistance and fitness. *Adv. Pharmacol.* **56**:257-96.
172. **Rambaut, A., D. Posada, K. A. Crandall, and E. C. Holmes.** 2004. The causes and consequences of HIV evolution. *Nat. Rev. Genet.* **5**:52-61.
173. **Ramirez, B. C., E. Simon-Loriere, R. Galetto, and M. Negroni.** 2008. Implications of recombination for HIV diversity. *Virus Res.* **134**:64-73.
174. **Requejo, H. I.** 2006. Worldwide molecular epidemiology of HIV. *Rev. Saúde Pública* **40**:331-345.
175. **Rhee, S. Y., T. F. Liu, M. Kiuchi, R. Zioni, R. J. Gifford, S. P. Holmes, and R. W. Shafer.** 2008. Natural variation of HIV-1 group M integrase: Implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**:74.
176. **Richman, D. D.** 2001. HIV chemotherapy. *Nature* **410**:995-1001.
177. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. *J. Mol. Evol.* **40**:249-259.
178. **Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. Mccutchan, S. Osmanov, M. Peeters, D. Pienniazek, M. Salminem, P. M. Sharp S. Wolisky, and B. Korber (ed.).** 1999. HIV-1 nomenclature proposal: A reference guide to HIV-1 classification. Los Alamos National Laboratory (disponível em

- <http://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/1999/6/nomenclature.pdf>).
179. **Roquebert, B., and A. G. Marcelin.** 2008. The involvement of HIV-1 RNase H in resistance to nucleoside analogues. *J. Antimicrob. Chemother.* **61**:973-975.
 180. **Roques, P., D. L. Robertson, S. Souquière, C. Apetrei, E. Nerrienet, F. Barré-Sinoussi, M. Muller-Trutwin, and F. Simon.** 2004. Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. *AIDS* **18**:1371-1381.
 181. **Rose, P. P., and B. Korber.** 2000. Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation. *Bioinformatics* **16**:400-401.
 182. **Ross, E. K., A. J. Buckler-White, A. B. Rabson, G. Englund, and M. A. Martin.** 1991. Contribution of NF-κB and Sp1 binding motifs to the replicative capacity of human immunodeficiency virus type 1: distinct patterns of viral growth are determined by T-cell types. *J. Virol.* **65**:4350-4358.
 183. **Rozanov, M., U. Plikat, C. Chappey, A. Kochergin, and T. Tatusova.** 2004. A web-based genotyping resource for viral sequences. *Nucleic Acids Res.* **32**:654-659.
 184. **Saitou, N., and N. Masatoshi.** 1987. The Neighbour-Joining method - a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406-425.
 185. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
 186. **Santiago, M., L., M. Lukasik, S. Kamenya, Y. Li, F. Bibollet-Ruche, E. Bailes, M. N. Muller, M. Emery, D. A. Goldenberg, J. S. Lwanga, A. Ayoub, E. Nerrienet, H. M. McClure, J. L. Heeney, D. P. Watts, A. E. Pusey, D. A. Collins, R. W. Wrangham, J. Goodall, J. F. Brookfield, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 2003 Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J. Virol.* **77**:7545-7562.
 187. **Santos Costa, Q., S. Coelho-Martins, and J. Azevedo Pereira.** 2007. The Prestige! Differences between HIV and SIV: virus-cell initial interaction-tropism and pathogenicity, p 197-212. *In* SIDAnet, Associação Lusófona (ed.), 8º Congresso Virtual: Novas perspectivas sobre a infecção VIH/SIDA e doenças relacionadas.
 188. **Sarafianos, S. G., B. Marchand, K. Das, D. M. Himmel, M. A. Parniak, S. H. Hughes, and E. Arnold.** 2009. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **385**:693-713.

189. **Sarngadharan, M. G., M. Popovic, L. Bruch, J. Schüpbach, and R. C. Gallo.** 1984. Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS. *Science* **224**:506-508.
190. **Schüpbach, J. M., R. Popovic, V. Gilden, M. A. Gonda, M. G. Sarngadharan, and R. C. Gallo.** 1984. Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* **224**:503-505.
191. **Semenova, E. A., C. Marchand, and Y. Pommier.** 2008. HIV-1 integrase inhibitors: Update and perspectives. *Adv. Pharmacol.* **56**:199-228.
192. **Serrao, E., S. Odde, K. Ramkumar, and N. Neamati.** 2009. Raltegravir, elvitegravir, and metoogravir: the birth of “me-too” HIV-1 integrase inhibitors. *Retrovirology* **6**:25.
193. **Shafer, R. W.** 2002. Genotypic testing for human immunodeficiency virus type 1 drug resistance. *Clin. Microbiol. Rev.* **15**:247-277.
194. **Shafer, R. W., and J. M. Schapiro.** 2008. HIV-1 Drug resistance mutations: an updated framework for the second decade of HAART. *AIDS Rev.* **10**:67-84.
195. **Shafer, R. W., K. Dupnik, M. A. Winters, and S. H. Eshleman.** 2001. HIV-1 reverse transcriptase and protease sequencing for drug resistance studies, p.83-133. *In* C. Kuiken, B. Foley, B. Hahn, P. Marx, F. Mccutchan, J. W. Mellors, S. Wolinsky and B. Korber (ed.), *HIV Sequence Compendium 2001*, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
196. **Sharp, P., M., D. L. Robertson, and B. H. Hahn.** 1995. Cross-species transmission and recombination of 'AIDS' viruses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **349**:41-47.
197. **Sharp, P., M., E. Bailes, R. R. Chaudhuri, C. M. Rodenburg, M. O. Santiago, and B. H. Hahn.** 2001. The origins of acquired immune deficiency syndrome viruses: where and when? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**:867-876.
198. **Sichtig, N., S. Sierra, R. Kaiser, M. Daumer, S. Reuter, E. Schulter, A. Altmann, G. Fatkenheuer, U. Dittmer, H. Pfister, and S. Esser.** 2009. Evolution of raltegravir resistance during therapy. *J. Antimicrob. Chemother.* **64**:25-32.
199. **Sierra, S., B. Kupfer, and R. Kaiser.** 2005. Basics of the virology of HIV-1 and its replication. *J. Clin. Virol.* **34**:233-244.
200. **Siliciano, R. F (ed.).** 2006. *Acquired immune deficiency syndrome (AIDS).* **Wiley, J.** Encyclopedia of Life Sciences. (disponível em <http://www3.interscience.wiley.com>).

201. **Simon, F, P. Maucière, P. Roques, I. Loussert-Ajaka, M. C. Müller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barré-Sinoussi, and F. Brun-Vézinet.** 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**:1032-1037.
202. **Sluis-Cremer, N., E. Kempner, and M. A. Parniak.** 2003. Structure–activity relationships in HIV-1 reverse transcriptase revealed by radiation target analysis. *Protein Sci.* **12**:2081-2086.
203. **Sluis-Cremer, N., N. A. Temiz, and I. Bahar.** 2004. Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. *Curr. HIV Res.* **2**:323-332.
204. **Spira, S., M. A. Wainberg, H. Loemba, D. Turner, and B. G. Brenner.** 2003. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J. Antimicrob. Chemother.* **51**:229-240.
205. **Stanford Database Group** (ed.). 1998-2010. Stanford University HIV Drug Resistance Database (disponível em <http://hivdb.stanford.edu/index.html>).
206. **Stebbing, J., B. Gazzard, and D. C. Douek.** 2004. Where does HIV live? *N. Engl. J. Med.* **50**:1872-1880.
207. **Stegen, K., E. Demecheleer, N. De Cabooter, D. Nges, M. Temmerman, P. Ndumbe, K. Mandaliya, J. Plum, and C. Verhofstede.** 2006. A sensitive in-house RT-PCR genotyping system for combined detection of plasma HIV-1 and assessment of drug resistance. *J. Virol. Methods* **133**:137-145.
208. **Suzuki, Y., and R. Craigie.** 2007. The road to chromatin-nuclear entry of retroviruses. *Nat. Rev. Microbiol.* **5**:187-196.
209. **Swanson, C. M., and M. H. Malim.** 2008. SnapShot: HIV-1 proteins. *Cell* **133**:742.
210. **Tachedjian, G., H. E. Aronson, M. de los Santos, J. Seehra, J. M. McCoy, S. P. Goff.** 2003. Role of residues in the tryptophan repeat motif for HIV-1 reverse transcriptase dimerization. *J. Mol. Biol.* **326**:381-396.
211. **Tachedjian, G., J. Radzio, and N. Sluis-Cremer.** 2005. Relationship between enzyme activity and dimeric structure of recombinant HIV-1 reverse transcriptase. *Proteins* **60**:5-13.
212. **Takebe, Y., R. Uenishi, and X. Li.** 2008. Global molecular epidemiology of HIV: Understanding the genesis of AIDS pandemic. *Adv. Pharmacol.* **56**:1-25.
213. **Takebe, Y., S. Kusagawa, and K. Motomura.** 2004. Molecular epidemiology of HIV: Tracking AIDS pandemic. *Pediatr. Int.* **46**:236-244.

214. **Tamura K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA 4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596-1599.
215. **Tapp Alter, K.** 2001. Protease inhibitors. *Science* **8**:59-64.
216. **Taylor, B. S., M. E. Sobieszczyk, F. E. McCutchan, and S. M. Hammer.** 2008. The challenge of HIV-1 subtype diversity. *N. Engl. J. Med.* **358**:1590-1602.
217. **Thali, M., A Bukovsky, E. Kondo, B. Rosenwirth, C. T. Walsh, J. Sodroski, and H. G. Göttinger.** 1994. Functional association of cyclophilin A with HIV-1 virions. *Nature.* **372**:363-365.
218. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673-4680.
219. **Thomson, M. M., and Rafael Nájera.** 2007. Increasing HIV-1 Genetic Diversity in Europe. *J Infect Dis.* **196**:1120-1124.
220. **Thomson, M. M., E. Delgado, I. Herrero, M. L. Villahermosa, E. Vazquez-de Parga, M. T. Cuevas, R. Carmona, L. Medrano, L. Perez-Alvarez, L. Cuevas, and R. Najera.** 2002. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. *J. Gen. Virol* **83**:107-119.
221. **Thomson, M. M., E. Delgado, N. A. Ocampo, M. L. Villahermosa, A. Marino, I. Herrero, M. T. Cuevas, E. Vazquez-de Parga, L. Perez- Alvarez, L. Medrano, J. A. Taboada, R. Nájera, and the Spanish Group for Antiretroviral Studies in Galicia.** 2001. HIV-1 genetic diversity in Galicia Spain: BG intersubtype recombinant viruses circulating among injecting drug users. *AIDS* **15**:509-516.
222. **Todd, S., C. G. Anderson, D. J. Jolly, and C. S. Craik.** 2000. HIV protease as a target for retrovirus vector-mediated gene therapy. *Biochim. Biophys. Acta.* **1477**:168-188.
223. **Tossonian, H. K., J. D. Raffa, J. Grebely, M. Viljoen, A. Mead, M. Khara, M. Mclean, A. Krishnamurthy, S. Devlaming, and B. Conway.** 2008. Primary drug resistance in antiretroviral-naïve injection drug users. *Int. J. Infect. Dis.* **13**:577-583.
224. **Uchil, P. D., and W. Mothes.** 2009. HIV Entry Revisited. *Cell* **137**:402-404.

225. **UNAIDS/WHO (ed.).** 2009. AIDS epidemic update. (disponível em <http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2009/default.asp>).
226. **Van Baelen, K., V. Van Eygen, E. Rondelez, and L. J. Stuyver.** 2008. Clade-specific HIV-1 integrase polymorphisms do not reduce raltegravir and elvitegravir phenotypic susceptibility. *AIDS* **22**:1877-1880.
227. **Van Heuverswyn, F., Y. Li, C. Neel, E. Bailes, B. F. Keele, W. Liu, S. Loul, C. Butel, F. Liegeois, Y. Bienvenue, E. M. Ngolle, P. M. Sharp, G. M. Shaw, E. Delaporte, B. H. Hahn, and M. Peeters.** 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature*. **444**:164.
228. **Van Maele, B., K. Busschots, L. Vandekerckhove, F. Christ, and Z. Debyser.** 2006. Cellular co-factors of HIV-1 integration. *Trends Biochem Sci.* **31**:98-105.
229. **Van Maele, B., and Z. Debyser.** 2005. HIV-1 integration: an interplay between HIV-1 integrase, cellular and viral proteins. *AIDS Rev.* **7**:26-43.
230. **Vandegraaff, N., and A. Engelman.** 2007. Molecular mechanisms of HIV integration and therapeutic intervention. *Expert Rev. Mol. Med.* **9**:1-19.
231. **Vasconcelos, C. A., J. M. Oliveira, J. P. Casquilho, J. A. Malhado, J. A. Vera L. F. Caldeira, M. M. Doroana and, R. A. Marques (ed.).** 2009. Recomendações portuguesas para o tratamento da infecção VIH/SIDA. Coordenação nacional para a infecção VIH/SIDA. (disponível em <http://www.sida.pt>).
232. **Vergne, L., L. Stuyver, M. Van Houtte, C. Butel, E. Delaporte, and M. Peeters.** 2006. Natural polymorphism in protease and reverse transcriptase genes and in vitro antiretroviral drug susceptibilities of non-B HIV-1 strains from treatment-naïve patients. *J. Clin. Virol.* **36**:43-49.
233. **Vergne, L., M. Peeters, E. Mpoudi-Ngole, A. Bourgeois, F. Liegeois, C. Toure-Kane, S. Mboup, C. Mulanga-Kabeya, E. Saman, J. Jourdan, J. Reynes, and E. Delaporte.** 2000. Genetic diversity of protease and reverse transcriptase sequences in non-subtype-B human immunodeficiency virus type 1 strains: evidence of many minor drug resistance mutations in treatment-naïve patients. *J. Clin. Microbiol.* **38**:3919-3925.
234. **Vergne, L., and M. Peeters.** 2007. The challenge of HIV sequence diversity in the envelope glycoprotein, p.17-29. *In* J. D. Reeves and C. A. Derdeyn (ed.), entry inhibitors in HIV therapy, Birkhäuser Basel, Montpellier.
235. **Vidal, N., S. E. Bazepeo, C. Mulanga, E. Delaporte, and M. Peeters.** 2009. Genetic characterization of eight full-length HIV type 1 genomes from the Democratic Republic of Congo (DRC) reveal a new subsubtype, A5, in the A radiation that predominates in the recombinant structure of CRF26_A5U. *AIDS Res. Hum. Retroviruses.* **25**:823-832.

236. **Vingerhoets J., M. Buelens, and M. Peeters.** 2007. Impact of baseline mutations on the virologic response to TMC125 in the phase III clinical trials DUET-1 and DUET-2, abstract 32. 11th European AIDS Conference European AIDS Clinical Society (EACS), Madrid.
237. **Wapling, J., K. L. Moore, S. Sonza, J. Mak, and G. Tachedjian.** 2005. Mutations that abrogate human immunodeficiency virus type 1 reverse transcriptase dimerization affect maturation of the reverse transcriptase heterodimer. *J. Virol.* **79**:10247-10257.
238. **Wiessing, L., M. J. Van de Laar, M. C. Donoghoe, B. Guarita, D. Klempová, and P. Griffiths (ed.).** 2008. HIV among injecting drug users in Europe: increasing trends in the East. *Euro Surveill.* (disponível em <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19067>).
239. **Wlodawer, A., M. Miller, M. Jaskólski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, and S. B. Kent.** 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science.* **245**:616-621.
240. **Wlodawer, A., and A. Gustchina.** 2000. Structural and biochemical studies of retroviral proteases. *Biochim. Biophys. Acta.* **1477**:16-34.
241. **Worobey, M., M. Gemmel, D. E. Teuwen, T. Haselkorn, K. Kunstman, M. Bunce, J. J. Muyembe, J. M. Kabongo, R. M. Kalengayi, and E. Van Marck.** 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960 *Nature* **455**:661-664.
242. **Yebra, G., M. de Mulder, J. del Romero, C. Rodríguez, and A. Holguín.** 2010. HIV-1 non-B subtypes: high transmitted NNRTI-resistance in Spain and impaired genotypic resistance interpretation due to variability. *Antiviral Res.* **85**:409-417.
243. **Yerly, S., B. Hirschel, C. Gaille, L. Kaiser, and L. Perrin.** 2007. Natural polymorphism of HIV-1 subtypes B and non-B integrase gene p. 129. *In* 14th Conference on Retroviruses and Opportunistic Infections, Los Angeles, CA.
244. **Zeinalipour-Loizidou, E., C. Nicolaou, A. Nicolaides, and L. G. Kostrikis.** 2007. HIV-1 integrase: From biology to chemotherapeutics. *Curr. HIV Res.* **5**:365-388.
245. **Zhang, M., B. Foley, A. K. Schultz, J. P. Macke, I. Bulla, M. Stanke, B. Morgenstern, B. Korber, and T. Leitner.** 2010. The role of recombination in the emergence of a complex and dynamic HIV epidemic. *Retrovirology* **7**:25.
246. **Zhu, T., B. T. Korber, A. J. Nahmiask, E. Hooper, P. M. Sharp, and D. D. Ho.** 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**:594-597.

6. Anexos

Anexo 2 - Composição dos meios e soluções utilizados

1. Meio LB:

10 g Bacto-triptona

5 g extracto de levedura

10 g NaCl

15 g agar*

Adicionar água para 1000 ml e autoclavar

* Não adicionar para LB líquido

2. Meio de criopreservação:

10% (v/v) DMSO

10% (v/v) glicerol

3. Solução de agarose (1,2%) (p/v) (50 ml):

0,6 g agarose

50 ml TAE 0,5X

4. Tampão Tris-acetato-EDTA (TAE) 50X:

242 g Tris-base

57,1 ml ácido acético glacial

100 ml 0,5 M EDTA, pH 8,0

Adicionar água para 1000 ml

5. Tampão de aplicação:

0,25% (p/v) azul de bromofenol

40% (p/v) sacarose

6. Tampão da T4 DNA ligase 2X (Promega, EUA):

60 mM Tris-HCl, pH 7,8

20 mM MgCl₂

20 mM DTT

2 mM ATP

10% (v/v) PEG

7. Meio de transformação e armazenamento (TSS):

LB 1X

10% (v/v) PEG6000

5% (v/v) DMSO

50 mM MgSO₄

(pH 6,5)

8. Placas LB agar/X-gal/IPTG/ampicilina:

LB sólido [agar Bacto a 1,5% (p/v) em LB líquido]

100 µg/ml ampicilina

40 µg/ml X-gal

0,17 mM IPTG

9. Tampão Tris-EDTA-glicose (TEG):

50 mM glicose

10 mM EDTA

25 mM Tris-HCl, pH 8,0

10. Solução de lise (SDS alcalino):

0,2 N NaOH

1% (p/v) SDS

11. Tampão da EcoRI 10X (Fermentas, Canadá):

50 mM Tris-HCl, pH 7,5

10 mM MgCl₂

100 mM NaCl

0,02% (v/v) Triton X-100

0,1 mg/ml BSA

Anexo 3 - Tabela contendo os dados clínico-epidemiológicos dos indivíduos (n=51) cujas amostras foram estudadas neste trabalho, estando indicado a origem e identificação da amostra (ID), bem como idade, sexo, ano de infecção e país de origem dos indivíduos, carga viral, origem dos parceiros sexuais, prática de relações sexuais não protegidas e de partilha de seringas, estágio da infecção e terapêutica anti-retroviral. Taipas- Centro de Atendimento a Toxicodependentes das Taipas; INSA- Instituto Nacional de Saúde Doutor Ricardo Jorge; HFF- Hospital Prof. Doutor Fernando Fonseca; F- Feminino; M- Masculino; CV- Carga viral; N- Não; S- Sim; FA- Fase assintomática; TAR- Terapêutica anti-retroviral; ?- Dado desconhecido.

Origem Amostras	ID	Idade (anos)	Sexo	Ano de Infecção	País de Origem	CV (cópias de RNA/ml)	Origem dos Parceiros Sexuais	Relações Sexuais Não Protegidas	Partilha de Seringas	Estádio da Infecção	TAR
Taipas	TP02	42	F	?	Portugal	?	Portuguesa	N	S	FA	N
Taipas	TP03	39	M	?	Portugal	?	Portuguesa	S	S	FA	S
Taipas	TP04	42	M	?	Portugal	?	?	S	S	FA	S
Taipas	TP08	37	M	?	Portugal	?	?	S	N	FA	S
Taipas	TP13	40	M	?	Portugal	?	Portuguesa	N	N	FA	N
Taipas	TP16	39	M	?	Portugal	?	Portuguesa	S	S	FA	N
Taipas	TP19	41	F	?	Portugal	?	Portuguesa	N	N	FA	S
Taipas	TP24	38	M	?	Portugal	?	Várias	N	N	FA	S
Taipas	TP38	36	M	?	Portugal	?	Portuguesa	N	N	FA	N
Taipas	TP42	44	M	?	Portugal	?	Várias	S	S	?	S
Taipas	TP44	28	F	?	Portugal	?	Portuguesa	N	N	?	?
Taipas	TP47	33	F	?	Portugal	?	Portuguesa	S	N	?	S

Taipas	TP55	37	F	?	Portugal	?	Portuguesa	S	S	?	S
Taipas	TP63	42	M	?	Angola	?	Várias	S	N	?	S
Taipas	TP71	46	F	?	Portugal	?	Portuguesa	S	N	?	?
Taipas	TP72	44	F	?	Portugal	?	Portuguesa	S	N	?	N
Taipas	TP85	32	M	?	Angola	?	Várias	S	S	?	N
Taipas	TP98	47	M	?	Portugal	?	Várias	S	S	?	N
Taipas	TP107	31	F	?	?	?	Portuguesa	S	N	?	?
Taipas	TP114	38	M	?	?	?	Portuguesa	S	N	?	?
INSA	TX08	34	M	1998	Portugal	6000	?	?	S	FA	?
INSA	TX12	38	M	1994	Portugal	22000	?	?	S	FA	?
INSA	TX13	32	M	2005	Moçambique	27000	?	?	S	FA	N
INSA	TX14	34	F	2005	Portugal	56000	?	?	S	FA	N
INSA	TX15	47	M	2005	Portugal	90000	?	?	S	FA	N

INSA	TX16	47	M	1989	?	35000	?	?	S	FA	?
INSA	TX18	36	M	2006	?	1900000	?	?	S	FA	N
INSA	TX20	38	M	1993	?	13000	?	?	S	FA	?
INSA	TX21	45	M	1996	?	11000	?	?	S	FA	?
INSA	TX23	32	F	2000	?	4200	?	?	S	FA	?
INSA	TX27	41	M	1986	Portugal	5300	?	?	S	FA	?
INSA	TX30	34	M	2006	?	1200	?	?	S	FA	N
INSA	TX32	41	F	1992	?	3900	?	?	S	FA	?
INSA	TX33	39	M	2005	?	1200	?	?	S	FA	N
INSA	TX34	37	M	1992	Portugal	15000	?	?	S	FA	?
INSA	TX37	24	M	2006	Portugal	160000	?	?	S	FA	N
INSA	TX40	30	M	1995	Marrocos	190000	?	?	S	FA	?
INSA	TX41	30	F	2005	Portugal	13000	?	?	S	FA	N

INSA	TX43	42	M	2006	Portugal	470	?	?	S	FA	N
INSA	TX45	37	M	1998	Portugal	43000	?	?	S	FA	?
INSA	TX46	26	F	2004	Cabo Verde	16000	?	?	S	FA	N
INSA	TX47	41	M	2005	Portugal	1700	?	?	S	FA	N
INSA	TX48	34	M	1994	EUA	610	?	?	S	FA	?
INSA	TX50	36	M	1999	Portugal	140000	?	?	S	FA	?
INSA	TX51	49	F	2005	Portugal	41000	?	?	S	FA	N
INSA	TX52	34	M	2005	Portugal	31000	?	?	S	FA	N
INSA	TX54	33	M	1998	Portugal	3000	?	?	S	FA	?
INSA	TX59	29	M	1995	Portugal	11000	?	?	S	FA	N
INSA	TX60	41	M	1996	Portugal	42000	?	?	S	FA	?
INSA	TX69	40	M	2005	Portugal	13000	?	?	S	FA	?
HFF	AS646	37	F	1998	Angola	?	?	?	?	?	N

Anexo 4 - Comparação dos resultados de genotipagem obtidos após construção das árvores filogenéticas para as três regiões codificantes individuais (PR, RT e IN) e para a sequência concatenada (PR+RT+IN).

	PR	RT	IN	Concatenada
TP02	B	B	B	B
TP03	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TP13	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TP16	U	B	B	B
TP19	CRF02_AG	G (02_AG)	CRF02_AG	CRF02_AG
TP24	B	B	B	B
TP38	B	B	B	B
TP47	*	*	G (14_BG)	---
TP71	B	B	B	B
TP72	G(14_BG)	G(14_BG)	G(14_BG)	G (14_BG)
TP85	*	G(14_BG)	*	---
TP98	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TP107	G(14_BG)/U	B	B	G(14_BG)/B
TP114	G(14_BG)/U	B	U/G(14_BG)/U	G(14_BG)/U/B/G(14_BG)
TX08	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TX12	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TX13	CRF02_AG	G(02_AG)	U	CRF02_AG/U
TX14	G(14_BG)	G(14_BG)/B	*	---
TX15	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TX16	B	B	G(14_BG)/U	B/G(14_BG)/U
TX18	G(14_BG)	G(14_BG)/B	G(14_BG)/U	G(14_BG)/B/G(14_BG)/U
TX21	B	B	B	B
TX23	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TX27	B	B	*	---
TX30	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TX32	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TX33	G(14_BG)	G(14_BG)	G (14_BG)	G(14_BG)
TX34	U	B	B	B
TX37	B	*	G(14_BG)	---
TX40	G(14_BG)	G(14_BG)/B	G(14_BG)	G(14_BG)/B/G(14_BG)
TX41	CRF02_AG	G(02_AG)	U	CRF02_AG/U
TX43	*	*	G(14_BG)/U	---
TX45	G(14_BG)	G(14_BG)	G(14_BG)	G (14_BG)
TX46	CRF02_AG	G(02_AG)	U/CRF02_AG/U	CRF02_AG/U
TX47	G(14_BG)/U	B	G(14_BG)/U	B/U/B/G(14_BG)
TX48	G(14_BG)	G(14_BG)	G(14_BG)	G(14_BG)
TX50	B	B	B	B
TX51	G(14_BG)	G(14_BG)/U	G(14_BG)/U	G(14_BG)/U/G(14_BG)
TX52	CRF02_AG	G(02_AG)	CRF02_AG/G(14_BG)	CRF02_AG/G(14_BG)
TX54	B/U	B	B	B
TX59	B	B	B	B
TX60	B	B	B	B

TX69	B	B	B		B
AS646	F (F1)	F (F1)	F (F1)		F (F1)

*** Região não amplificada ou não editada.**

--- Sequência não utilizada para a construção da árvore filogenética, por falta de, pelo menos, uma das regiões codificantes em estudo.

Notas: (1) Nesta tabela, as regiões não genotipáveis (U) são consideradas apenas quando correspondem a um mínimo de, aproximadamente, 20 % do tamanho do fragmento analisado, sendo este critério também aplicado para as regiões de genótipo definido [PR > 60 nucleótidos; RT > 120 nucleótidos; IN > 160 nucleótidos; sequência concatenada (PR+RT+IN) > 300 nucleótidos];

(2) As sequências assinaladas a vermelho são recombinantes intergenotípicos, relativamente às sequências concatenadas em estudo.

Anexo 5 - Código de uma letra dos aminoácidos

A	Alanina
C	Cisteína
D	Ácido aspártico
E	Ácido glutâmico
F	Fenilalanina
G	Glicina
H	Histidina
I	Isoleucina
K	Lisina
L	Leucina
M	Metionina
N	Asparagina
P	Prolina
Q	Glutamina
R	Arginina
S	Serina
T	Treonina
V	Valina
W	Triptofano
Y	Tirosina

Anexo 6a – Polimorfismos genéticos não associados a resistência presentes na região da protease de sequências dos subtipos B (n=12) e não-B (n=29).

Mutação	B	não-B	<i>p</i> *	Mutação	B	não-B	<i>p</i> *
V11L	0	1	1,000	Y59F	0	1	1,000
T12A	0	1	1,000	D60E	1	0	1,000
T12E	1	0	0,310	Q61E	1	0	1,000
T12K	1	1	0,505	Q61N	0	1	1,000
T12P	0	1	1,000	I62V	2	1	0,200
T12S	3	1	0,068	L63F	1	0	1,000
I13V	4	26	0,001	L63L	1	1	0,505
K14E	1	0	0,31	L63P	8	2	0,000
K14R	3	16	0,098	L63S	0	1	1,000
K14T	0	1	1,000	I64M	0	6	0,151
I15V	7	5	0,020	I64V	3	0	0,021
G16E	0	2	1,000	C67E	0	1	1,000
G17D	0	1	1,000	C67S	0	1	1,000
G17E	0	5	0,298	C67Y	0	1	1,000
L19I	3	1	0,068	H69K	0	21	0,000
K20I	0	24	0,000	H69R	0	1	1,000
K20R	4	0	0,005	K70G	1	0	1,000
E21K	0	1	1,000	K70Q	0	1	1,000
E35A	0	1	1,000	K70R	0	5	0,298
E35D	0	22	0,000	I72R	0	1	1,000
M36I	0	26	0,000	I72V	3	1	0,068
M36L	0	1	1,000	V75I	0	1	1,000
N37E	3	0	0,021	V77I	6	1	0,001
N37K	0	1	1,000	V82I	0	17	0,000
N37S	1	0	0,310	R87K	1	1	0,505
R41K	6	26	0,011	L89I	0	1	1,000
V56I	0	1	1,000	L89M	0	22	0,000
R57K	0	4	0,302				

*Os valores de *p* assinalados a vermelho correspondem às assinaturas do subtipo B, enquanto que os assinalados a azul correspondem às assinaturas dos subtipos não-B.

Anexo 6b – Polimorfismos genéticos não associados a resistência presentes na região da transcriptase reversa de sequências dos subtipos B (n=17) e não-B (n=24).

Mutação	B	não-B	p*	Mutação	B	não-B	p*
I31L	1	0	0,415	K173A	0	1	1,000
I31V	1	0	0,415	K173I	0	9	0,005
V35I	2	1	0,560	K173L	0	1	1,000
V35K	0	2	0,502	K173Q	1	0	0,415
V35L	1	0	0,415	K173R	0	1	1,000
V35T	0	21	0,000	K173T	0	9	0,005
E36D	0	4	0,128	K173V	0	1	1,000
T39A	0	2	0,502	Q174E	0	1	1,000
T39L	0	1	1,000	Q174K	0	19	0,000
E40D	0	1	1,000	Q174R	0	1	1,000
K43R	0	9	0,005	P176S	0	1	1,000
K49R	1	2	1,000	D177E	4	19	0,001
E53D	0	3	0,254	I178L	0	3	0,254
V60I	1	19	0,000	I178M	0	5	0,065
S68G	0	1	1,000	I178V	1	1	1,000
R83K	10	0	0,000	V179I	1	0	0,415
D86E	0	1	1,000	I195L	0	1	1,000
A98S	0	16	0,000	G196E	7	2	0,021
K102Q	1	0	0,415	Q197K	1	0	0,415
K102R	1	1	1,000	T200A	2	19	0,000
D121Y	0	2	0,502	T200I	1	0	0,415
K122E	0	17	0,000	T200V	0	1	1,000
K122R	0	1	1,000	K201R	0	1	1,000
D123E	9	1	0,001	I202V	2	1	0,560
D123N	0	3	0,254	E203D	0	1	1,000
D123S	0	13	0,000	E204D	0	2	0,502
R125I	0	1	1,000	E204G	0	2	0,502
T131S	0	1	1,000	E204K	1	3	0,629
I132V	0	2	0,502	R206K	0	1	1,000
I135K	0	1	1,000	Q207A	0	5	0,065
I135L	1	0	0,415	Q207E	0	7	0,029

I135T	7	11	1,000	Q207G	0	1	1,000
I135V	2	6	0,433	Q207K	1	7	0,110
I142V	0	2	0,502	Q207N	0	1	1,000
I159V	1	0	0,415	Q207S	0	1	1,000
S162A	0	5	0,065	Q207T	0	1	1,000
S162C	1	2	1,000	R211K	12	7	0,012
S162H	0	1	1,000	R211N	1	1	1,000
S162N	0	1	1,000	R211Q	0	3	0,254
S162R	0	1	1,000	R211S	0	11	0,001
S162Y	1	0	0,415	F214E	1	0	0,415
T165I	1	0	0,415	F214L	1	2	1,000
K166R	3	2	0,633	P217T	1	0	0,415
E169D	0	1	1,000				

*Os valores de p assinalados a vermelho correspondem às assinaturas do subtipo B, enquanto que os assinalados a azul correspondem às assinaturas dos subtipos não-B.

Anexo 6c – Polimorfismos genéticos não associados a resistência presentes na região da integrase de sequências dos subtipos B ($n=13$) e não-B ($n=28$).

Mutação	B	não-B	p^*	Mutação	B	não-B	p^*
K7E	1	0	0,317	G134N	0	25	0,000
K7I	0	1	1,000	I135V	1	5	0,645
K7Q	0	1	1,000	K136D	1	0	0,317
E11D	7	7	0,089	K136I	0	1	1,000
K14R	1	15	0,006	K136Q	1	1	0,317
S17C	0	1	1,000	K136T	0	26	0,000
S17N	0	7	0,077	K156R	0	1	1,000
S17T	0	2	1,000	G163D	1	0	0,317
R20G	0	1	1,000	G163E	0	1	1,000
R20K	5	6	0,280	V165A	0	1	1,000
A21T	1	2	1,000	V165I	1	2	1,000
S24N	0	1	1,000	R166K	0	1	1,000
L28I	1	0	0,317	D167G	0	1	1,000
P30S	0	1	1,000	H171Q	0	1	1,000
V31I	0	7	0,077	K173R	0	2	1,000

V31M	1	0	0,317	F181L	0	2	1,000
V32I	7	3	0,005	K188R	0	1	1,000
I36V	0	2	1,000	G189R	0	2	1,000
V37I	1	0	0,317	G189W	0	1	1,000
S39C	7	0	0,000	G193E	2	2	0,579
D41N	1	0	0,317	G193R	2	1	0,232
L45I	1	0	0,317	V201I	13	25	0,539
E48K	0	1	1,000	A205G	0	1	1,000
M50I	3	5	0,692	A205S	1	0	0,317
K71Q	0	2	1,000	A205T	0	1	1,000
I72T	0	1	1,000	T206S	1	20	0,000
I72V	7	4	0,019	K211E	0	1	1,000
L74I	1	0	0,317	K211R	1	0	0,317
I84L	1	1	0,539	K215N	4	0	0,007
P90S	1	0	0,317	T218I	0	2	1,000
P90T	1	1	0,539	I220V	0	1	1,000
A91T	1	0	0,317	N222H	1	0	0,317
S93C	1	0	0,317	N222K	1	0	0,317
E96D	1	0	0,317	Y227F	0	15	0,001
Y99F	0	1	1,000	D229V	0	1	1,000
F100C	0	1	1,000	D232E	2	0	0,095
L101I	4	28	0,000	D232G	0	1	1,000
L101V	1	0	0,317	L234I	1	21	0,000
L104V	0	1	1,000	L234V	2	7	0,692
G106A	1	0	0,317	G247R	0	2	1,000
K111R	1	0	0,317	V250A	1	0	0,317
T112A	2	0	0,095	I251L	1	0	0,317
T112I	2	14	0,045	D253E	1	0	0,317
T112M	0	4	0,288	D253Y	0	1	1,000
T112V	0	4	0,288	N254G	1	0	0,317
I113L	0	1	1,000	N254K	0	1	1,000
I113V	1	0	0,317	S255N	0	19	0,000
S119P	0	6	0,152	D256E	1	18	0,001

S119R	1	3	0,317	I257T	1	0	0,317
S119T	0	3	0,539	I257V	0	1	1,000
T122I	0	3	0,539	R262K	0	2	1,000
T124A	2	26	0,000	I267G	0	1	1,000
T124G	0	1	1,000	I268L	0	3	0,539
T124N	6	0	0,000	I268M	1	0	0,317
T124Q	0	1	1,000	R269E	1	0	0,317
T125A	6	25	0,005	R269G	1	0	0,317
T125V	2	1	0,232	R269K	0	4	0,288
K127R	0	2	1,000	R269Q	1	0	0,317
G134D	0	2	1,000	D270H	0	2	1,000

*Os valores de p assinalados a vermelho correspondem às assinaturas do subtipo B, enquanto que os assinalados a azul correspondem às assinaturas dos subtipos não-B.

Anexo 6d – Polimorfismos genéticos não associados a resistência presentes nas regiões da protease, transcriptase reversa e integrase de sequências dos subtipos não-G e G.

Protease				Transcriptase reversa				Integrase			
Mutação	não-G (n=24)	G (n=17)	p^*	Mutação	não-G (n=28)	G (n=13)	p^*	Mutação	não-G (n=29)	G (n=12)	p^*
I13V	13	17	0,001	V35T	10	11	0,006	K14R	8	8	0,034
K20I	7	17	0,000	K43R	2	7	0,002	L101I	20	12	0,039
E35D	8	14	0,004	V60I	7	13	0,000	T112I	5	11	0,000
M36I	9	17	0,000	A98S	8	13	0,000	T124A	16	12	0,007
R41K	16	16	0,056	K122E	6	11	0,000	T125A	19	12	0,020
H69K	4	17	0,000	K173I	0	9	0,000	G134N	13	12	0,009
V82I	2	15	0,000	K173T	5	4	0,429	K136T	14	12	0,001
L89M	5	17	0,000	Q174K	6	13	0,000	T206S	9	12	0,000
				D177E	10	13	0,000	Y227F	3	12	0,000
				T200A	8	13	0,000	L234I	10	12	0,000
				Q207E	5	2	1,000	S255N	7	12	0,000
				R211S	0	11	0,000	D256E	7	12	0,000

*Os valores de p assinalados a vermelho correspondem às assinaturas do subtipo G.

Anexo 7 – Comparação dos resultados de genotipagem para as três regiões codificantes do gene *pol*, obtidos, neste trabalho após construção de árvores filogenéticas, com dados análogos para os genes *gag* e *env* gentilmente fornecidos por E. Pádua para as amostras com origem no INSA.

	<i>gag</i>	<i>pol</i>			<i>env</i>
	?	Protease	Transcritase Reversa	Integrase	gp120 (ansa V3)
TX08	G	G(14_BG)	G(14_BG)	G(14_BG)	G
TX12	CRF14_BG/B	G(14_BG)	G(14_BG)	G(14_BG)	B
TX13	A	CRF02_AG	G(02_AG)	U	A
TX14	G	G(14_BG)	G(14_BG)/B	*	G
TX15	CRF02_AG/CRF14_BG	G(14_BG)	G(14_BG)	G(14_BG)	A
TX16	B	B	B	G(14_BG)/U	B
TX18	B	G(14_BG)	G(14_BG)/B	G(14_BG)/U	B
TX21	B	B	B	B	B
TX23	G	G(14_BG)	G(14_BG)	G(14_BG)	G
TX27	B	B	B	*	B
TX30	G	G(14_BG)	G(14_BG)	G(14_BG)	B
TX32	G	G(14_BG)	G(14_BG)	G(14_BG)	B
TX33	G	G(14_BG)	G(14_BG)	G (14_BG)	G
TX34	B	U	B	B	B
TX37	B	B	*	G(14_BG)	A (A1)
TX40	B	G(14_BG)	G(14_BG)/B	G(14_BG)	B
TX41	A (A1)	CRF02_AG	G(02_AG)	U	A
TX43	B	*	*	G(14_BG)/U	B

	<i>gag</i>	<i>pol</i>			<i>env</i>
	?	Protease	Transcritase Reversa	Integrase	gp120 (ansa V3)
TX45	G	G(14_BG)	G(14_BG)	G(14_BG)	B
TX46	A	CRF02_AG	G(02_AG)	U/CRF02_AG/U	A
TX47	G	G(14_BG)/U	B	G(14_BG)/U	B
TX48	G	G(14_BG)	G(14_BG)	G(14_BG)	G
TX50	B	B	B	B	B
TX51	B/J	G(14_BG)	G(14_BG)/U	G(14_BG)/U	A
TX52	A	CRF02_AG	G(02_AG)	CRF02_AG/G(14_BG)	A
TX54	B	B/U	B	B	B
TX59	B	B	B	B	B
TX60	B	B	B	B	B
TX69	B	B	B	B	B

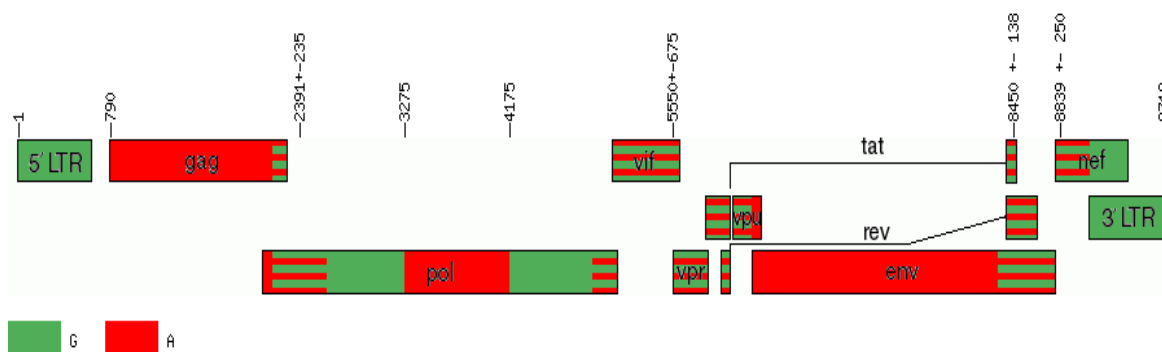
 - Região não amplificada ou não editada

? - Região do genoma desconhecida

Anexo 8 - Pontos de recombinação teóricos de formas recombinantes circulantes relevantes para este trabalho (figuras com origem em <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

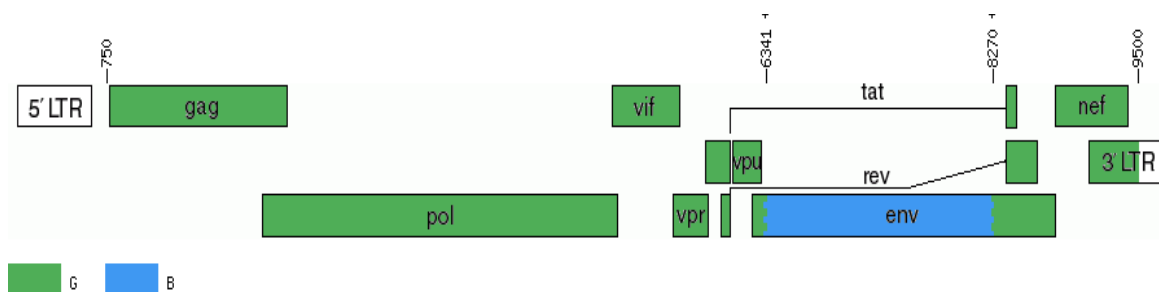
CRF02_AG

Pontos de recombinação teóricos: [1 a 789-G]; [790 a 2155-A]; [2156 a 2625-A/G]; [2626 a 3274-G]; [3275 a 4174-A]; [4175 a 4874-G]; [4875 a 6224-A/G]; [6225 a 8311-A]; [8312 a 8588-A/G]; [8589 a 9089 A/G]; [9090 a 9719-G].



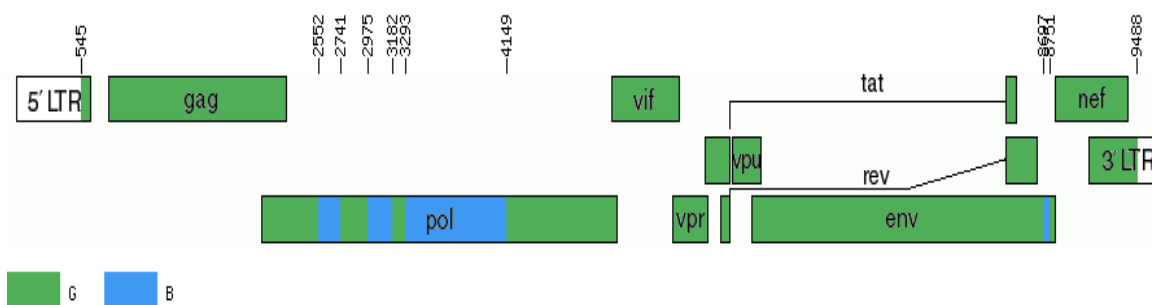
CRF14_BG

Pontos de recombinação teóricos: [750 a 6322-G]; [6323 a 6359-B/G]; [6360 a 8260-B]; [8261 a 8279-B/G]; [8280 a 9500-G].



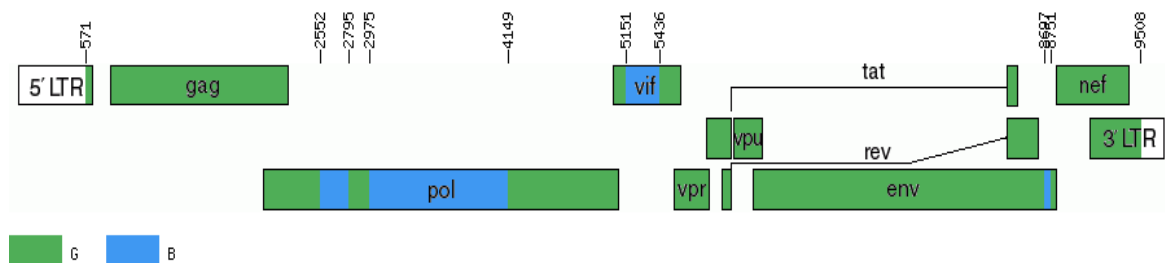
CRF20_BG

Pontos de recombinação teóricos: [545 a 2551-G]; [2552 a 2740-B]; [2741 a 2974-G]; [2975 a 3181-B]; [3182 a 3292-G]; [3293 a 4148-B]; [4149 a 8696-G]; [8697 a 8750-B]; [8751 a 9488-G].



CRF23_BG

Pontos de recombinação teóricos: [571 a 2551-G]; [2552 a 2794-B]; [2795 a 2974-G]; [2975 a 4148-B]; [4149 a 5150-G]; [5151 a 5435-B]; [5436 a 8696-G]; [8697 a 8750-B]; [8751 a 9508-G].



CRF24_BG

Pontos de recombinação teóricos: [571 a 2551-G]; [2552 a 2842-B]; [2843 a 2974-G]; [2975 a 3292-B]; [3293 a 3379-G]; [3380 a 4148-B]; [4149 a 8696-G]; [8697 a 8750-B]; [8751 a 9508-G].

